Appl. Math. Inf. Sci. **7**, No. 3, 1157-1163 (2013)

Applied Mathematics & Information Sciences
An International Journal

Dynamically Real-time Anomaly Detection Algorithm with Immune Negative Selection

Lingxi Peng[1], *Wenbin Chen*[1], *Dongqing Xie*[1], *Ying Gao*[1] *and Chunlin Liang*[2,*]

[1] Department of Computer and Education software, Guangzhou Univ., Guangzhou 510006, China
[2] School of Information, Guangdong Ocean University, Zhanjiang 524088, China

Received: 20 Sep. 2012, Revised: 28 Jan. 2013, Accepted: 12 Feb. 2013
Published online: 1 May 2013

Abstract: Network anomaly detection has become the promising aspect of intrusion detection. The existing anomaly detection models depict the detection profiles with a static way, which lack good adaptability and interoperability. Furthermore, the detection rate is low, so they are difficult to be deployed the realtime detection under the high-speed network environment. In this paper, the excellent mechanisms of self-learning and adaptability in the human immune system are referred and a dynamic anomaly detection algorithm with immune negative selection, named as *DADAI*, is proposed. The concepts and formal definitions of antigen, antibody, and memory cells in the network security domain are given; the dynamic clonal principle of antibody is integrated; the mechanism of immune vaccination is discussed, and the dynamic evolvement formulations of detection profiles are established (including the detection profiles' dynamic generation and extinction, dynamic learning, dynamic transformation, and dynamic self-organization), which will accomplish that the detection profiles dynamically synchronize with the real network environment. Both our theoretical analysis and experimental results show that *DADAI* is a good solution to network anomaly detection, which increase the veracity and timeliness on anomaly detection problem.

Keywords: Artificial immune, network anomaly detection, negative selection, intelligent system.

1. Introduction

Network anomaly detection systems has become the promising aspect of network intrusion detection systems. The network anomaly detection [1] techniques have been discussing and a variety of detection methods had been proposed. From the technical point of view, these methods can be divided into three mainly categories. The first is statistical analysis-based network anomaly detection methods including the threshold detection, GLR (Generalized Likelihood Ratio) test, exponential smoothing technique, Markov model [2], autoregressive [3], D-S evidence theory [4], wavelet analysis, Bayesian clustering, etc [5]. The biggest advantage of these methods is that they can learn the users' habits, and have high detection rate and availability. However, it is difficult to determine the threshold value, and attackers can train the abnormal profiles into normal ones. The second is data mining-based network anomaly detection methods including the classification, association rules, clustering [6], decision tree, etc. The biggest characteristic of these methods is that they can automatically construct the repository, which improves the detection accuracy. However, with the continuous updating of repository, behavior characteristics will be more complicated. Furthermore, as time goes by, the repository becomes more and more large, which increases the detection time. The third is machine learning-based network anomaly detection methods including the neural networks, genetic algorithms, fuzzy rules, reinforcement learning [7] as well as online adaptive network anomaly detection [8].

These methods still have a lot of problems to be solved including the system's incompleteness on theory and practicality. There is actually a direct analogy between the computer network security and the biological immune system (BIS) in a human body. Both of them have to maintain stability in a changing environment. The network security techniques based on artificial immune system (AIS) have the features of diversity, self-adaptation and robustness. Thus, they are considered a very promising research direction in network security [9,10].

* Corresponding author e-mail: flyingday@139.com

According to the negative selection algorithm, Dagupta et al. proposed an immune-based network anomaly detection method and applied on the benchmark intrusion detection dataset-Lincoln dataset, which confirm the feasibility of this method [11]. Seredynski et al. also tested the immune system-based network anomaly detection method using the Lincoln data sets, and the experimental results show that the system can detect all the abnormal network scenarios [12]. However, there are three mainly defects for the above anomaly immune-based detection methods. The first is that the detection profiles are described as a static way and will not change after the definition. The static definition of detection profiles can not well adapt to the real complex network environment, which will cause the system's high false positives and false negative rate. The other problem is the detection profiles will become complex and huge as time goes by, thus the system lose the ability to learn. In addition, most of the existing anomaly detection models are complicated, the detection range is not comprehensively enough, and detection efficiency is low, so they are difficult to be applied on large-scale high-speed real-time network anomaly detection problem. Furthermore, different anomaly detection systems and even every parts in the one system lack the effective mechanism on information sharing and collaboration. These issues have become the major obstacle for the immune-based network anomaly detection in realities. In order to solve these problems, Zhou J. et al. proposed a dynamic radius of self-element combining with statistical methods to control the number of detector [13]. Sun et al. put forward the dynamic anomaly trigger mechanism on immune detection profiles [14], which achieves the dynamical detector updating and memory association. These works try to solve above two problems. However, these research only consider dynamic detector update under static network environment (the experiments take the static dataset), which is difficult to adapt to the complex practical dynamic network environment. Furthermore, these works do not consider the real-time anomaly detection under large-scale high-speed network environment including the anomaly detection systems' cooperativity, so there are still long distances on specific applications.

In this paper, we steal the excellent learning adaptive mechanism of human immune system and propose a dynamically real-time anomaly detection algorithm with immune negative selection algorithm, referred as *DADAI*. The definitions and description of antigen, antibody and memory cells are given in network environment. Then, the dynamic clonal antibody theory is combined with *DADAI*. Furthermore, the immune vaccination and vaccine distribution mechanisms are discussed, and the dynamic evolution equations are established, which will accomplish that the detection profiles dynamically synchronize with the real network environment. The simulation experiments were carried, and the comparative results show that the *DADAI* improves the accuracy and timeliness compared with existing network anomaly detection methods.

The rest of the paper is organized as follows. In Section 2 we present our theoretical model. In Section 3, simulations and experimental results are provided. Related works and comparison are given in Section 4. Finally, Section 5 contains our summary and conclusions.

2. Proposed algorithm

The algorithm is contained in Figure 1. As shown in the figure, the antigen is defined as $AG =< d, c >$, where d is the gene, $d \subset D$, $D = [0,1]^l (l > 0)$, and c is species including *Normal* and *Abnormal*. The IP packets transmitted on the network will be presented and summarized, which will form the network transmission speed, the number of real-time connection session, the unit number of TCP-SYN packets, etc. These characteristics data describing network transactions will be standardized and form the gene of antigen, $Ag.d$. The immune cells are defined as $IC =< d, c, age, count >$, where d is the gene, $d \in D$, c is species including *Normal* and *Abnormal*, *age* is life, $age \in N$, *count* is the match number of with the detected antigen, $count \in N$, N is natural number. The immune cells are made up of immune cells MC and antibody cell AB, where $IC = MC \bigcup AB$ and $MC \bigcap AB = \emptyset$. f_{match} is defined as the match function between immune cells and antigens, which can be calculated by Euclidean, Manhattan, etc. The smaller of the match value, the closer of these two immune cells, and vice versa.

2.1. The dynamic evolution of detection profiles

The detection profile description in existing network anomaly detection methods or technology is a static way, which seldom changes after the definition. Thus, they can not well adapt to actual network environment. It's the important reason that there are high false positive rates and false negative rates in these anomaly detection systems. In addition, the static description of detection profiles lack good self-learning, self-adaptability, and self-organizing ability, which seriously influence the efficiency of anomaly detection and can not meet the requirement of real network monitoring.

In our algorithm, the memory cells at time t, $MC(t)$, which represent abnormal profiles. It is depicted as following *Eq.* (1).

$$MC(t) = \begin{cases} MC(0) & t = 0 \\ MC_{young}(t) \cup MC_{doat}(t) \cup \{ag\}(t) & \\ \cup \{mc\}(t) \cup MC_{vaccine}(t) - MC_{var}(t) & \\ -MC_{dead}(t) - MC_{merge}(t) & t > 0 \end{cases} \quad (1)$$

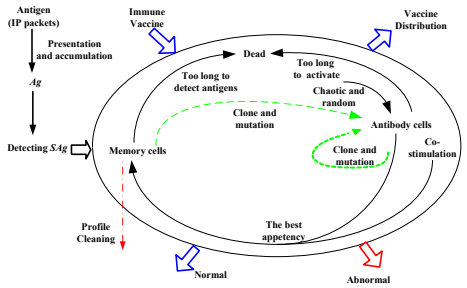

Figure 1: Dynamically real-time anomaly detection with immune negative selection.

$$MC(0) = \begin{cases} \Phi & iff \text{ no initial dataset} \\ \text{identified data after the random} & \\ \text{generation and collection} & else \end{cases} \quad (2)$$

$$f_{costimulation}(x) = \begin{cases} 0 \text{ normal profiles} \\ 1 \text{ abnormal profiles} \end{cases} \quad (3)$$

The initial set of memory cells, $MC(0)$, can be null, or identified data after the random generation and collecting, which is as *Eq.* (2) depicts. If the memory cells representing the abnormal profiles match the antigen, and the co-stimulation $f_{costimulation}(x)$ (the external system's input signal from external interface, 0 is normal, 1 is abnormal) is also abnormal, then the species will not change, which indicates that the abnormal network activity has not changed.

$$MC_{young}(t) = \{x | x.d = y.d, x.age = 0, x.c = y.c, x.count = y.count + 1, \forall y \in MC(t-1) \wedge < y.d, ag.d > \in Match \wedge f_{costimulation}(x) = 1\} \quad (4)$$

$$MC_{vaccine\_distribuate}(t) = \left\{ x \left| \begin{array}{l} x \in MC(t-1) \\ \wedge x.count > \gamma \end{array} \right. \right\} \quad (5)$$

$$MC_{doat}(t) = \{x | \forall y \in MC(t-1) \wedge < y.d, ag.d > \notin Match \wedge f_{costimulation}(x) = 1, x.d = y.d, x.c = y.c, x.age = y.age + 1\} \quad (6)$$

$$\{ag\}(t) = \{x | \forall y \in MC(t-1) \wedge < y.d, ag.d > \notin Match \wedge f_{costimulation}(ag) = 1, x.d = age.d, x.c = 1, x.age = 0\} \quad (7)$$

$$\{mc\}(t) = \{x \left| \arg\min_{y \in AB} f_{match}(y, ag) < \arg\min_{z \in MC} f_{match}(z, ag) \wedge f_{costimulation}(y) = 1, x.d = y.d, x.age = 0, x.count = 1\} \right. \quad (8)$$

$$MC_{var}(t) = \{x | \forall y \in MC(t-1) \wedge < y.d, ag.d > \in Match \wedge f_{costimulation}(x) = 0, x.d = y.d, x.age = 0, x.count = 0\} \quad (9)$$

$$MC_{dead}(t) = \{x | x \in MC(t-1) \wedge x.age > \beta\} \quad (10)$$

$$MC_{merge}(t) = \{x | \forall y \forall z \in MC(t-1) \wedge (z.d, y.d) \in Match, y.d \neq z.d, f_{match}(y, ag) > f_{match}(z, ag), x.age = z.age, x.d = z.d, x.count = z.count\} \quad (11)$$

Eq. (1) gives the constitution of abnormal profiles $MC(t)$ at time t. Specifically, at time t-1, *Eq.* (4) shows that if the abnormal profiles, $MC_{young}(t)$, match the detecting antigen and the and the $f_{costimulation}(x)$ is abnormal, which indicates that the profiles match the detecting antigen, the count number will be added 1. If the count number is larger than γ, the profiles will be distributed to other machines, which is shown as *Eq.* (5).

Memory cells have the lifecycle β. If memory cells detect the intrusion antigen, the age will be set as 0. Otherwise the age will be added 1 as *Eq.* (6) depicts. The setting of age is to limit the number of memory cells. If the age is larger than β, the memory cells will be removed from the normal profiles. $MC_{doat(t)}$ represents the profiles that do not match the detecting antigen. If there is no any profile matches the antigen, the antigen will be directly joined into the memory cells. which is $ag(t)$ in Eq. (7). The process is similar with the antigen presenting, which is to collect new profiles elements from the internet. In addition, if the match value between some antibody and detecting antigen is smaller than the match value between any memory cell and detecting antigen, the antibody will be added into the normal profiles if the $f_{costimulation}(x)$ is abnormal. The process is shown in *Eq.* (8), which is $mc(t)$. *DADAI* chooses the antigens and joins them into the abnormal profiles, which indicate that our model can learn from the new antigens and carry out the first immune reply. The memory of these antigens will be saved, which will enhance the model's self-learning and adaptability.

The implanted profiles, $\mathrm{MC}_{vaccine}(t)$, will also be consolidated. There are two ways for the vaccination, the first is from the other machines, and the second is manual implantation by the system administrator. Through the implanted vaccines, the search processes can be accelerated, and the learning time can be reduced, which will improve the interoperability of the anomaly detection systems. In *Eq.* (9), the variation profiles will be removed from profiles, $MC_{var}(t)$, which indicates that profiles match the detecting antigen with the normal $f_{costimulation}(x)$. In this case, these profiles have varied. If the profiles have not detected the antigen for a long time and age is too old, $MC_{dead}(t)$ describing dying memory cells will be removed from the profiles, which is shown in

Eq. (10). $MC_{dead}(t)$ are profiles that the system does not need any more and should be removed, which will help to control the size of memory cells and improve the detection speed. With the continual learning, the memory cells will be joined into the memory cells set. In order to control the number of the memory cells, if the two match memory cells match each other, the memory cell with the larger match value will be removed from the memory cells set, which is $MC_{merge}(t)$ as *Eq*. (11) describing. In this way, *DADAI* adjusts the memory cell and control the number of memory cell, which help to improve the model's self-organization and self-maintenance.

Through its own immune surveillance, *DADAI* will remove the mutated profiles at any moment, and deduce the false positive. The false positive is to take the normal network activity as the abnormal network behavior, which will avoid the false alarm rate. The false positive is to take the normal network activity as the abnormal network behavior, which will avoid the false alarm rate. For another, *DADAI* dynamically increase the abnormal profiles elements, which will increase the description scope and deduce the false negative. The false negative is to take the abnormal network activity as the normal network behavior. Combining with the following dynamic antibody clone and elimination model, *DADAI* can overcome the high false alarm rate in the existing anomaly detection methods, and increase the model's self-adaptability. We propose a new dynamic description of detection profiles, which naturally collect the new elements from the Internet and set the evolution cycle of profile. Meantime, the useless profile elements will be eliminated and the dynamic evolution concepts are introduced, which subtly unify the variability and invariability. Thus, the dynamical detection profiles will synchronously evolve with the real network world. *DADAI* can detect the unknown pattern with good self-adaptability. In this way, *DADAI* ensures the system's stability in macro and avoid the indefinitely growth of memory cells.

2.2. Dynamic antibody clone and elimination

$$AB(t) = \begin{cases} AB(0) & t = 0 \\ AB(t-1) \cup AB_{clone}(t-1) - \\ AB_{dead}(t-1) \cup AB_{chaotic\_new}(t-1) & t > 0 \end{cases} \tag{12}$$

$$AB(0) = \begin{cases} \Phi & iff \text{ no colllection dataset} \\ \text{collected identified data} & else \end{cases} \tag{13}$$

$$K_{n+1} = \begin{cases} K_n/\lambda & 0 < K_n < \lambda \\ (1.0-K_n)/(1.0-\lambda) & \lambda < K_n < 1.0 \end{cases} \tag{14}$$

$$AB_{clone}(t) = AB_{atom}(t) \cup AB_{mtoa}(t) \tag{15}$$

$$AB_{atom}(t) = \bigcup_{x \in \{mc\}(t)} \bigcup_{i=1}^{\left\lceil hR \bullet cR \bullet \frac{1}{f_{match}(x,ag)} \right\rceil} \{x_i^{'}\}, \quad x_i^{'}.age = 0, x_i^{'}.count = 0, x_i^{'}.d = f_{variation}(x.d) \tag{16}$$

$$AB_{mtoa}(t) = \bigcup_{x \in MC_{young}(t)} \bigcup_{i=1}^{\left\lceil hR \bullet cR \bullet \frac{1}{f_{match}(x,ag)} \right\rceil} \{x_i^{'}\}, \quad x_i^{'}.age = 0, x_i^{'}.count = 0, x_i^{'}.d = f_{variation}(x.d) \tag{17}$$

$$AB_{dead}(t-1) = \{x \,|\, x \in AB(t-1) \wedge x.age > \alpha\} \tag{18}$$

The detection profiles in existing network anomaly detection methods are usually directly generated, and there are few studies on candidate detection profiles. Thus, they lack the ability to detect similar variant attacks. The system's self-learning and self-adaptability needs to be improved. Let the antibody set, *AB*, represent candidate detection profiles. In this paper, we propose a method of antibodies' dynamic clone and elimination.

The antibody set $AB(t)$ at time t, is the antibody set $AB(t\text{-}1)$ unites the clone antibody set $AB_{clone}(t-1)$, and minus the dead antibody set $AB_{dead}(t-1)$, unites the new antibody set $AB_{chaotic\_new}(t-1)$ generated by Kent chaotic map (a very good uniformity of Chaos), which is shown as *Eq*. (12). *Eq*. (13) shows the antibody set at the initial time, which indicates $AB(0)$ can be null or some collected and identified data set. *Eq*. (14) shows how the gene of antibody can be generated where $\lambda \in (0,1)$ and K_0 can be randomly set between 0 and 1.

Eq. (15) shows that the antibody's clone to antigen, which can be divided into two kinds of cases. The first is when the antibodies evolve into memory cells, $AB_{atom}(t)$. The second is when the memory cells match the detecting antigen, $AB_{mtoa}(t)$. They are shown in *Eq*. (16) and Eq. (17), respectively. The purpose of cloning is to produce more immune cells to detect similar attacks. $AB_{atom}(t)$ is the clone of first reply, which needs to go through a long learning period so the efficiency is relatively low. $AB_{mtoa}(t)$ is the clone of second reply, which is the cloning of memory cells. *cR* and *hR* are integer set by user. *cR* will determine the number of clonal antibody cell and the typical value is 10. *hR* is positive constant, which generally takes the positive integer. After the clone, the gene of antibodies will be mutated. The mutation probability is the match value between antibody and detecting antigen. Finally, the mutated clone antibodies will be added into the antibody set of *AB*. Compared with the first reply, the second reply is rapid and strong, which needs no learning process. Once the antigen matches memory cells, the response will be immediate. If the antibodies do not evolve into memory cells at number of lifecycle, which will be eliminated as *Eq*. (18) depicts.

DADAI implements the antibodies and memory cells' dynamic cloning and mutation, which enhances the

model's self-adaptive learning, and improves the detection efficiency to variant abnormal profiles. In this way, *DADAI* will have strong detection abilities to detect distributed denial of service attack.

Compared with the existing research, the proposed algorithm adopts the Kent chaos mapping to randomly generate the antibodies. Furthermore, the antibodies' dynamic clone and elimination will assure the diversity and adaptability of antibody. *DADAI* has the parallelism and robustness of immune system with the chaos' initial value sensitivity, ergodicity, randomness and other characteristics, which meantime overcomes the low search efficiency in the immune algorithm. In this way, *DADAI* has better search efficiency and global search optimization ability to detect more unknown patterns.

2.3. Network anomaly detection process

The traditional detection models need to define the outline and abnormal value of normal detection profile, which itself is a difficult problem. In our model, the real-time network anomaly detection process does not need to give the outline of normal pattern. The detection is determined by the memory cells. In *Eq*. (19), the specie of detecting antigen will depend on the matching memory cells.

$$ag.c = \begin{cases} normal \quad else \\ abnormal \; iff \; exists \; y \in MC(t) \wedge \\ < y.d, ag.d > \in Match \end{cases} \tag{19}$$

From the detection process, we can see that the spatiotemporal complexity of detection has linear relationship with the number of memory cells and antibodies. In addition, the lifecycle of antibody cells and memory cells, α and β, can be set larger, respectively. However, at the evolution stage of detection profiles, they can be set smaller to further control the number of antibody cells and memory cells and enhance the adaptability.

3. Experiments

The used Breast Cancer Wisconsin dataset in this study is taken from UCI machine learning repository [15]. This dataset contains 30 real-valued input features, which contains 569 samples belonging to two different classes (357 "benign" (normal) cases, 212 "malignant" (abnormal) cases).

For test results to be more valuable, k-fold cross-validation is used among the researchers. It minimizes the bias associated with the random sampling of the training. In this method, whole data is randomly divided to k mutually exclusive and approximately equal size subsets. The detection paradigm trained and tested k times. We used this method as 10-fold cross-validation in our experiments.

In this study, the detection accuracy for the dataset was measured according to *Eq*. (20) and *Eq*. (21), where T is the set of data items to be detected (the test set), $t \in T$, $t.c$ is the class of the item t, and detect(t) returns the detection of t by *DADAI*.

$$\text{accuracy}(T) = \frac{\sum_{i=1}^{|T|} \text{assess}(t_i)}{|T|}, t_i \in T \tag{20}$$

$$\text{assess}(t) = \begin{cases} 1, \text{ if detect}(t) = t.c \\ 0, \text{ otherwise} \end{cases} \tag{21}$$

Besides of Detection Accuracy, Sensitivity, Specificity and Average detection rate (ADR) measures are also given in two-class problems as in Eq. (22)-(24), respectively, where TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{22}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{23}$$

$$\text{ADR} = (\text{Sensitivity} + \text{Specificity})/2 \tag{24}$$

Authors	Method	Accuracy/%
Our study	*DADAI*	97.2
Wojnarski M.	3-NN standard Manhattan	97.1
Wojnarski M.	kNN with DVDM distance	97.1
Ster B. et al.	21-NN standard Euclidean	96.9
Ster B. et al.	LVQ	96.6
Ster B. et al.	kNN, Euclidean/Manhattan	96.6
Hamilton H.J., et al.	NB – naive Bayes	96.4
Wojnarski M.	C4.5 (decision tree)	96.0
Hamilton H.J., et al.	Assistant R tree (ASR)	94.7

Table 1: Detection accuracies of *DADAI* with detection accuracies obtained by other methods in the literature

The obtained detection accuracy is 97.2% using 10-fold cross-validation over the dataset. Also, sensitivity, specificity and ADR values for the dataset were obtained as 96.2%, 97.8% and 97.0%, respectively. The obtained detection accuracy by combine of *DADAI* for the dataset is the highest method among methods report from the literature. In view of detection accuracy, Table 1 [16] is shown to compare these results with *DADAI*.

Network anomaly detection models	Dynamic detection	Unknown attacks	Accuracy	Network anomaly detection ability
Statistics-based methods [1–5]			relatively high	It is difficult to determine the threshold value, and the attackers can train the abnormal patterns into normal ones.
Data mining-based methods [6]		√	high	These methods can automatically construct the repository, but the repository is more and more larger as time goes.
Machine learning-based methods [7,8]		√	high	The completeness of theory and the practicality of the system have to be solved.
Existing immune-based methods [10,11]		√	common	The detection profiles are described in a static way, thus these methods lose the ability to learn.
DADAI	√	√	relatively high	The detectors synchronously evolve with the real network environment. *DADAI* has good self-learning, adaptability, self-organization and interoperability, which can be used for real-time large data processing under high-speed network environment.

Table 2: The Comparison of *DADAI* and related works.

4. Related works

In actual network environment, the external network environment is always complex and changing, so we have to dynamically depict the abnomal detection profiles to adapt the network environment. In this paper, a novel dynamic immune-based network anomaly detection algorithm, referred as *DADAI*, is proposed. The detection profiles of *DADAI* can synchronously evolve with the network environment. Table 2 shows the comparison of *DADAI* and related studies.

From Table 2, we can see that the existing anomaly detection models describe the detection profiles in a static way, which lack good self-learning, self-organization, adaptability, and interoperability. Thus, they are difficult to be efficiently deployed in large-scale high-speed network environment. *DADAI* novelly integrates the latest research achievements of life science into network security, and realizes that the detectors synchronously evolve with the network environment. In this way, *DADAI* effectively resolves the problem that the static detection profiles have poor interoperability and are hard to be deployed in large-scale network environment. Specifically, as a new anomaly detection method, *DADAI* has good self-learning, adaptability, self-organization and interoperability; secondly, at the stages of learning and detecting, *DADAI* scan data only once, so the detecting speed is fast and efficiency is high, which can be deployed in large-scale high-speed network environment; finally, *DADAI* model can globally extract and interoperate the detection information, so it has strong collaboration capabilities.

5. Conclusion

In this paper, a novel immune-based dynamic anomaly detection algorithm with negative selection, *DADAI*, is proposed. Both our theoretical analysis and experimental results prove that *DADAI* has good self-learning, adaptability, self-organization and interoperability. Thus, *DADAI* can be effectively deployed on the real-time network intrusion detection under high-speed network environment, which will help to improve accuracy and timeliness of abnormal events. When the large-scale security incidents arise, *DADAI* can conduct quick and efficient monitoring. In addition, *DADAI* is also a generic anomaly detection model. The dynamic realtime detection principle can be also widely used on search engine, credit card fraud detection, speech and handwriting recognition research, and stock market analysis, etc.

Acknowledgement

The authors acknowledge the financial support of the National Natural Science Foundation of China under Grant No. 61100150 and No. 11271097, and the Natural Science Foundation of Guangdong Province of China under Grant No. S2011040004528, No.S2011040004121 and No. S2011040003843.

References

[1] D. Lee and D. Won, Applied Mathematics & Information Sciences **6**, 209 (2012).
[2] J. J. Flores, A.Antolino and J. M.Garcia, Proc. Sixth International Conference on Networking and Services (ICNS), 271 (2010).
[3] B. X. Zou, Chin. J. Comput. **26**, 940 (2003).
[4] J. W. Zhuge, D. W. Wang, Y. Chen, Z. Y. Ye and W. Zou, Chin. J. Softw. **17**, 463 (2006).
[5] Y.X. Mao, Applied Mathematics & Information Sciences **5**, 97 (2011).
[6] Y. J. Zhou, C. Xu and J. G. Li, Chin. J. Commu. **31**, 18 (2010).
[7] S. X. Wu and W. Banzhaf, Appl. Soft Comput.**10**, 1 (2010).

[8] J. W. Zhuge, D. W. Wang, Y. Chen, Z. Y. Ye and W. Zou, J. Comput. Res. Dev. **47**, 485 (2010).
[9] W. Chen, X.J. Liu, T. Li., Y.Q. Shi, X.F. Zheng and H. Zhao, Int. J. Comput. Intell. Syst. **4**, 410 (2011).
[10] X.C Zhao, G.L. Liu, H.Q. Liu, G.S. Zhao and S.Z. Niu, Int. J. Comput. Intell. Syst. **3(S1)**, 1 (2010).
[11] D. Dasgupta and F. Gonzalez, IEEE T. on Evol. Comput. **6**, 281 (2002).
[12] F. Seredynski and P. Bouvry, Comput. Commun. **30**, 740 (2007).
[13] J. Zhou and D. Dasgupta, Inform. Sciences. **179**, 1390 (2009).
[14] F.X. Sun and T.S. Huang, Chin. J. Comput. **29**, 463 (2006).
[15] C. L. Blake and C. J.Merz, UCI Repository of Machine Learning Databases. http://www.ics.uci.edu/ mlearn/ (2012).
[16] D. Wlodzislaw, Datasets used for classification: comparison of results. http://www.is.umk.pl/projects/datasets.html (2012).

Lingxi Peng obtained his Ph.D. in computer science from Sichuan University (China) in 2008. He is currently an associate professor at the College of Computer Science and Educational Software, Guangzhou University. His research interests include artificial immune, algorithm design and analysis, bioinformatics algorithms, database, network security, graph algorithms, graph mining, computational complexity, etc.

Wenbin Chen received his M.S. degree in mathematics from Institute of Software, Chinese Academy of Science in 2003, and the Ph.D. degree in computer science from North Carolina State University, U.S.A in 2010. He is currently an associate professor at the College of Computer Science and Educational Software, Guangzhou University. His research interests include algorithm design and analysis, bioinformatics algorithms, graph algorithms, graph mining, computational complexity, database, etc.

Dongqing Xie received his bachelor degree in applied mathematics from Xidian University in 1985, the M.S. degree in software engingering from Xidian University in 1988, China, and the Ph.D. degree in computer science from Hunan University in 1999, China. He is currently a professor at the College of Computer Science and Educational Software, Guangzhou University. His research interests include algorithm design and analysis, network security, graph algorithms, graph mining, applied mathematics, etc.

Ying Gao obtained his Ph.D. in computer science from South China University of Technology (China). He is currently a professor at the College of Computer Science and Educational Software, Guangzhou University. His research interests include algorithm design and analysis, blind signal separation, signal processing, bioinformatics algorithms, graph algorithms, graph mining, computational complexity, etc.

Chunlin Liang received his M.S. degree in software engineering from Institute of Software, South China University of Technology in 2005. He is currently a lecturer at the School of Information, Guangdong Ocean University. His research interests include algorithm design and analysis, bioinformatics algorithms, graph algorithms, graph mining, computational complexity, etc.