International Journal of Multimedia and Ubiquitous Engineering
Vol. 1, No. 3, September, 2006

Normal model and BPNN-Based Immunization of Anti-Worm Web System

Tao Gong[1], and **Zi-xing Cai**[2]
[1]College of Information Science and Engineering, Central South University
e-mail : taogongchina@gmail.com
[2]College of Information Science and Engineering, Central South University
e-mail : zxcai@ieee.org

Abstract Detection, recognition and learning of unknown worms has become a bottleneck of network security since a lot of variants of old worms and new worms occurred. To overcome this bottleneck, many traditional approaches were tested, but failed to detect all worms and recognize most unknown worms. In this paper, a normal model of a web system was proposed to detect all selfs and all non-selfs, especially all unknown worms. The normal model was based on the space attribute and time attribute of computer system, and a theorem proved validity of the normal model of the web system. Moreover, the web system was immunized through using a BP neural network (BPNN) as an adaptive learning mechanism. The learning mechanism was designed to recognize most unknown worms through trained BPNN, which was trained with the feature data of all known worms in the worm database. Besides, the innate non-self selection was utilized to recognize all known worms. Experiments validated effectiveness of this approach on the BPNN and the normal model.

Keyword: Normal model, Immunization, Worm, BP neural network, Web system

1. Introduction

More and more worms are dangerous to many networks, such as the Internet, though the early worm was designed for harnessing the (possible heterogenous) nodes of a multi-computer (networked computers) [1]. Staniford et al. described a worm that could infect the entire Internet in about 30 s [2]. A worm of this scale and speed could bring the entire network to a halt, or worse [3]. To prevent the spread and damage of the worms, some network techniques were used to detect and stop the propagation of the worms [4], [5], [6]. But most of these approaches failed to detect and recognize unknown worms, because the variants of old worms and new worms occurred continuously. The detection and recognition of unknown worms become a bottleneck of this problem. To overcome the bottleneck, an adaptive immunization approach was proposed for a web system on a BP neural network (BPNN) and the normal model of the web system.

Part of the reason of the BP neural network was the effectiveness of learning un-known objects. Though this approach took long time to train the neural network and the algorithm was easy to obtain its partial optimization, the BP network was fit to learn optimal parameter offline and some adaptive factors were added into the BP algorithm to improve its optimization.

On the other hand, part of the reason of the normal model was the full detection of all selfs of the web system and its non-selfs. Inspired from the natural immune system, self/non-self discrimination of the web system was built on its normal model, which was represented with the space-time features of its all normal components as similar to the special molecule structures of all normal cells of the body. All the 2-dimension features were represented and stored in the self database. Thus, all normal components were found in the self database. In other words, if an antigen didn't match any record in the self database, then the antigen was regarded as a non-self, perhaps a worm, and every non-self was detected through the normal model.

2. Normal Model of Web System

Suppose a web system S is comprised of 2 web directories and 70 HTML page files, then the file set of the system is represented as such.

$$\{p_{ij} \Big| \sum_{i=1}^{2} n_i = 70, p_{ij} \in S, j = 1,2,\cdots,n_i, i \in \{1,2\}\} \quad (1)$$

Here, p_{ij} denotes the jth file in the ith directory of the system S, and n_i denotes the sum of all files in the ith directory of the system S.

Thus, the normal states of the web system are represented with the normal states of all components in the web system. Inspired from the normal states of human immune system, the normal states of the web system are represented on the space and time dimensions of its normal components, to identify its selfs. Space dimension d_i of a component p_i is presented as its absolute pathname. Time dimension t_i of the component p_i is presented as its last revision time.

$$< D,T >= \{(d_i,t_i) | d_i \in p_i, t_i \in p_i, i = 1,2,\cdots,n\} \quad (2)$$

Here, n denotes the sum of all the normal components of the web system S.

Part of the reason of the 2-dimension modeling is the time-space positioning thought of Einstein [7]. In today's cyberspace, every component can also be uniquely identified with both its space dimension and its time dimension. If and only if every normal component of the normal web system is identified uniquely, the normal state of the system can be identified uniquely.

"This work was supported by two grants No.(60404021, 60234030) from National Natural Science Foundation of China."

Theorem 1: Suppose a web system S is comprised of n components, each of which has its unique absolute pathname and unique last revision time. Let the absolute pathname of a component p_i be d_{p_i} and the last revision time of the component p_i be t_{p_i}. Also suppose the cyberspace has the same order of space and time as nature and human society. Thus, the absolute pathnames and last revision time of all normal components in the normal system uniquely identify the normal states of the system.

[Proof] In the web system there are n components of web pages, which uniquely identify the system. In other words, if the components are all normal, then the web system is normal; but if any component is abnormal, then the web system is abnormal. Therefore, the normal state of every component in the web system should be identified with its space feature and time feature before the whole web system is identified. (1) Given an absolute pathname d_{p_i}, its corresponding component p_i is unique. Further given the last revision time t_{p_i} of the component p_i, its corresponding state is also unique. Moreover, the last revision time t_{p_i} is acquired when the web system is normal, i.e. the component p_i is also normal at that time. Therefore, the absolute pathname d_{p_i} and the last revision time t_{p_i} identify the normal state of the unique component p_i in both the space and time dimensions. (2) Given a component p_i of the system S, the absolute pathname d_{p_i} of the component is unique, because this fact is required by the rules of file management on any operation system. Besides, according to the hypothesis of this theorem, the last revision time t_{p_i} of the component is also unique, because the last revision time of everything is unique in the real world. Therefore, the absolute pathname and last revision time of every normal component in the normal web system uniquely identify the normal state of the component. Based on the unique identification between the web system and its components, the normal states of the web system are uniquely identified through the absolute pathname and last revision time of its each normal component. Hence, Theorem 1 is proved right.

Thus, all the normal states of the web system are represented as the above set of 2-element group, and the data of the model can be visualized as some molecules in a kind of 3-dimension space. On the Java platform, the absolute pathname and the last revision time of the component can be read through two methods of the class File. Modeling the normal states of the web system in 2 dimensions is shown in Fig. 1. Every component of the normal web system has its space property and time property, i.e. its absolute pathname and last revision time. The two properties can be read through Java method in the Java class package File, and the two properties of every component are also encapsulated as an immune object. The immune object is utilized in the immune logic when the web system is abnormal. Besides, the immune object is stored and accessed in the self database. The reason of modeling the normal states of the web system is the good test-bed of the immune model [8]. Thus, it is every fit for building its normal model and artificial immune system with pure Java, which is easy for developing web applications.

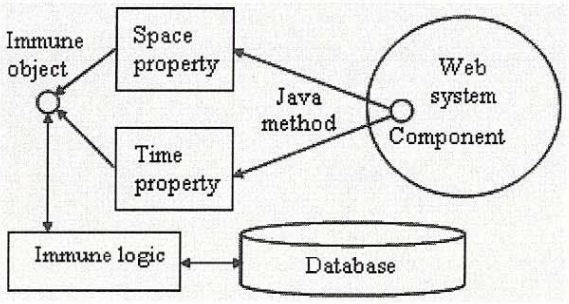


Fig. 1. Modeling normal states of the web system

3. BPNN-Based Immunization of Web System

In the natural immune system, adaptive learning of immune cells against unknown viruses is a kind of very complex process, which is even known little by doctors and immunologists. Alike, the adaptive learning of the artificial immune system against unknown worms is also very complex and is still a hard problem. To explore the secret, a BP neural network is built and used for the adaptive learning, and then the web system is immunized on its normal model and the BPNN.

The BP neural network consists of three tiers, i.e. input tier, hidden tier and output tier. In the input tier, 6 features of the known worms are represented as $\{x_i | i = 1,2,\cdots,6\}$, and these features include filename, coding language, proliferation manner, engine, feature string and damage. In the hidden tier, 3 types of known worms are represented as $\{x_i | i = 7,8,9\}$. In the output tier, 3 elimination schemas of known worms are represented as $\{x_i | i = 10,11,12\}$. With pure Java, the BP network is constructed as shown in Fig. 2.

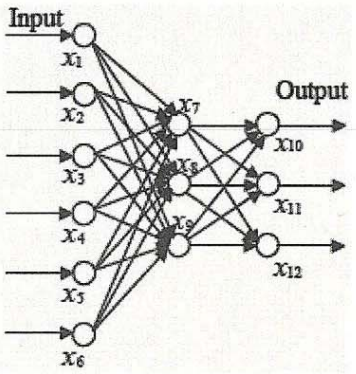


Fig. 2. Structure of BP neural network

The input data are acquired from all known worms as the samples of the BPNN in the worm database, which is used to represent and store the features of all known worms, such as the love worm, the happy-time worm and the code-red worm etc. The three types of worms have different elimination schemas, which are also represented and stored in the worm database. The BP neural network satisfies the following formulas.

$$P = \sum_y (\sum_z (d_{yz} - O_{yz})^2)$$
$$\Delta W_{i \to j} = r O_j (1 - O_j) \beta_j \quad (3)$$
$$\beta_j = \sum_k W_{j \to k} O_k (1 - O_k) \beta_k$$
$$\varepsilon_z = d_z - O_z$$

Here, P represents the performance of the BP neural network; y represents the training input; z represents the output node; d_{yz} represents the anticipant output of the node z through the input y, and d_j represents the anticipant output of the jth node; O_{yz} represents the actual output of the node z through the input y, and O_j represents the actual output of the jth node; $W_{i \to j}$ represents the weight value between the nodes of the ith tier and those of the jth tier, and $\Delta W_{i \to j}$ represents its change; r represents the learning rate; β_j represents the value of the jth node, ε_z represents the error of the output node.

The BP neural network is trained for 2117 times with the input data of all known worms, and the learning result is shown in Fig. 3. All known worms are used to train the BP neural network, so that the trained BPNN can recognize unknown worms from the non-selfs. The error of the output node approximates to zero as the BPNN is trained recursively.

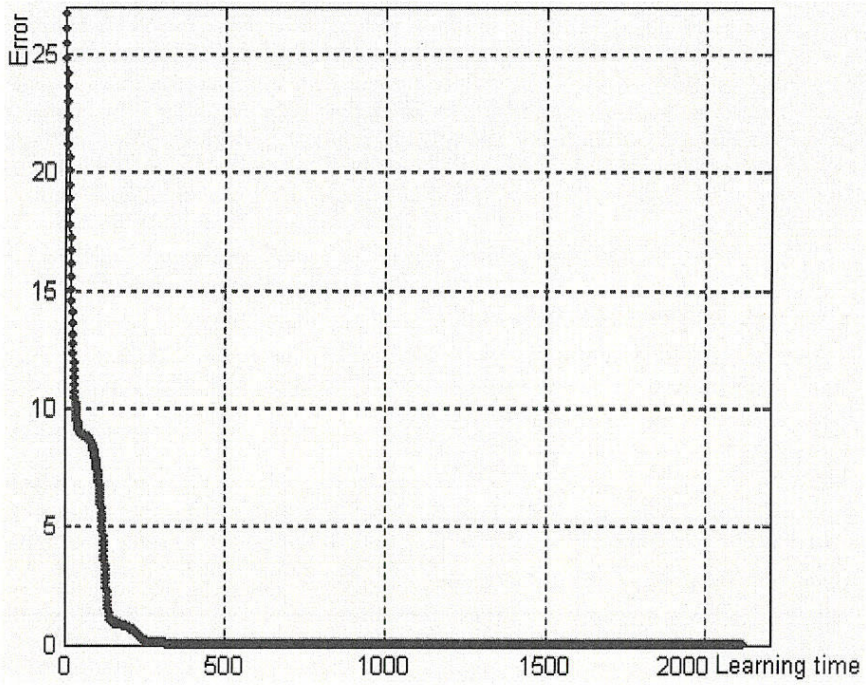


Fig. 3. Adaptive learning result of BP neural network

The adaptive learning of the BP neural network is improved though controlling the momentum of the learning rate, and used to recognize learn and memorize unknown worms from the unknown non-selfs. Besides, all known worms can be recognized through querying the features of the non-self in the worm database, which is the main function of the innate non-self selection.

After the BP network is trained for 2117 times through learning the samples of all known worms, the web system is tested to be infected by some love worms, happy-time worms and the variants of the love worms. Through 260 experiments of immunizing the web system, the detection rate of these worms is 100%, the recognition rate of love worms and happy-time worms is 100%, and the recognition rate of the unknown worms, i.e. the variants of the love worms, is 98%.

After some non-selfs are recognized known worms and unknown worms in the web system, the worms can be eliminated with some available means and other non-selfs are deleted directly, as visualized in Fig. 4. Black molecules represent the non-selfs and the gray molecules represent the normal components of the web system. The result figure shows that the non-selfs decrease in the web system. Furthermore, the infected components of the web system are repaired through its normal model and self database, after they are eliminated through the anti-worm immunization. At last, the whole web system is repaired back to its normal state and can work normally again. In Fig. 4, from left to right shows a process of immunization and failover in the web system. The ability of recognizing unknown worms lies on the types and amount of all known worms in the worm database. When some totally unknown worms are tested, the recognition rate may be much lower than 98%, but the recognition rate can be increased through using much more known worms as the samples for learning in the worm database.

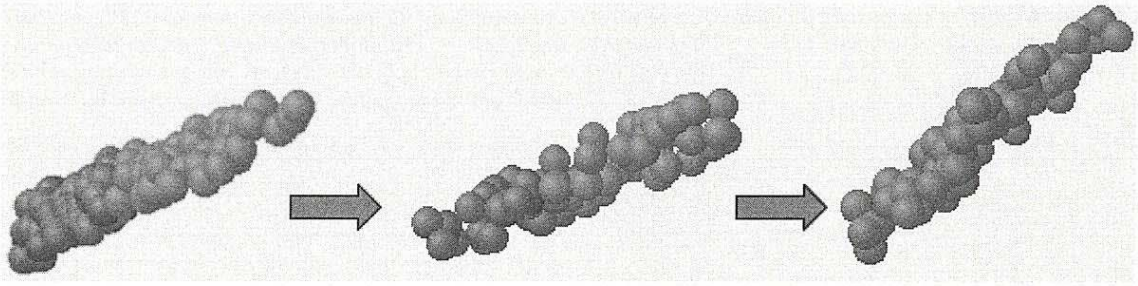


Fig. 4. Adaptive learning result of BP neural network

4. Conclusions

Though detection and recognition of unknown worms are very difficult for traditional approaches in the web system, the normal model is proved to identify the normal states of the system and validated its effective detection through repeated experiments. Also, the BP neural network is designed to learn unknown worms and validated its effective recognition through repeated experiments. Therefore, the normal model and the neural learning are both important and useful for immunizing the web system. Besides, the normal model is useful to repair the damaged web system.

Acknowledgements

Thanks for support of National Natural Science Foundation of China (60404021 & 60234030) and Excellent Doctoral Degree Project of Central South Univ. (040125).

References

[1] Y. Paker, T. Kindberg, "The worm program model: an application centred point of view for distributed architecture design," In: Proceedings of the 3rd workshop on ACM SIGOPS European workshop, pp. 1-4, 1988.

[2] S. Staniford, V. Paxson, N. Weaver, "How to own the Internet in your spare time," In: Proceedings of the USENIX Security Symposium USENIX Association, Berkeley, CA, pp. 149-167, 2002.

[3] J. Balthrop, S. Forrest, M. E. J. Newman, et al. "Technological Networks and the Spread of Computer Viruses," Science, Vol. 304, Issue 5670, pp. 527-529, 2004.

[4] E. Levy, "Worm propagation and generic attacks," IEEE Security and Privacy, Vol. 3, Issue 2, pp. 63-65, 2005.

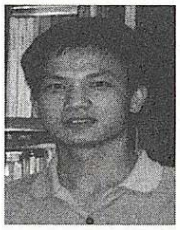
[5] R. S. Gray, V. H. Berk, "Rapid detection of worms using ICMP-T3 analysis," In: Proceedings of SPIE - The International Society for Optical Engineering, Orlando, pp. 89-101, 2004.

[6] C. C. Zou, W. Gong, D. Towsley, "Code Red Worm Propagation Modeling and Analysis," In: Proceedings of the 9th ACM conference on Computer and communications security, pp. 138-147, 2002.

[7] Einstein A. Relativity: the Special and General Theory, New York: Three Rivers Press, 1920.

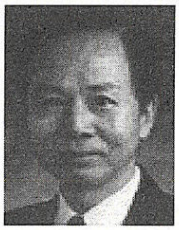
[8] T. Gong, Z. X. Cai, "An Immune Agent for Web-based AI Course," International Journal on E-Learning, Vol. 5, No. 4, pp. 493-506, 2006.

Authors



Tao Gong

Received a B.S. degree in applied mathematics from Hunan Science and Technology University, China, 2000, and M.S. degree in Computer Science from Central South University, China, 2003. Since 2003, he has studied for Ph D. degree in Computer Application Technology in Central South University, China. In 2004, he became the director of NSFC project on artificial immune system in Central South University. His research interests include Multimedia systems, Network Security, Immune Computation, Natural Computation, Artificial Intelligence. He is a Full Member of Sigma Xi Honor Society.



Zi-xing Cai

Received a B.S. degree in Electronic Engineering from Xi'an Jiaotong University, China, 1962. He is a Member of New York Academy, IEEE Senior Member, and Chair Professor at College of Information Science and Engineering, Central South University, China. His research interests include Artificial Intelligence, Intelligent Control and Intelligent Robots.