An improved Algorithm of Generating Network Intrusion Detector

Li Ma[1, a], Yan Chen[1, b]

[1]Department of Information Engineering Guilin University of Aerospace Technology, Guilin China

[a]E-mail: woshiml@sina.com, [b]E-mail: chen1982yan@163.com

Keywords: intrusion detection; negative algorithm; abnormal detection; optimization of Genelib

Abstract. On the basis of the analysis of the current algorithms on generating detector, an algorithm of generating network intrusion detector is proposed to avoid the disadvantages of negative selection. The algorithm takes advantage of the strategies such as gene library (GeneLib) etc to improve the efficiency and TP (true positive). Meanwhile, with the introducing of collaborative signals, the algorithm is good for recognizing the mistake operations so as to reduce FP (false positive). The experiment results show that the proposed algorithm is practical and efficient.

Introduction

Though increasing development and widely used of network brings people more and more convenience, it also brings many security problems to people's life and work. Intrusion detection system（IDS）is the special part of the security measures, so people have paid more and more attention to it.

Through the research on intrusion detection system and biological immune system, we can see that they are similar. The main goal of intrusion detection and immune system is to identify normal data and eliminate abnormal data. Hence, immune theory has good use for reference of network intrusion system. It is antibody that detects abnormal cells in immune system, so, the key to build the network intrusion system based on immune theory is to simulate gene library evolution and negative selection to generate intrusion detectors.

This paper analyzes the advantages and disadvantages of present algorithms for generating detectors, and then provides an improved algorithm of generating network intrusion detector (IDANS) on the basis of negative selection and clonal selection.

Traditional Algorithm of Generating Detectors

After the negative selection algorithm (NSA) was proposed by Forrest in 1994, negative selection becomes the necessary part of producing mature detectors in the applications of human immune system. NSA generates candidate detectors randomly and then matches them with "self" set. The candidate detectors will be deleted if one of them matches the "self" set, otherwise they will be converted to mature detectors. Repeat the steps until enough mature detectors are generated. The training and detecting stages of detectors are as Fig. 1.

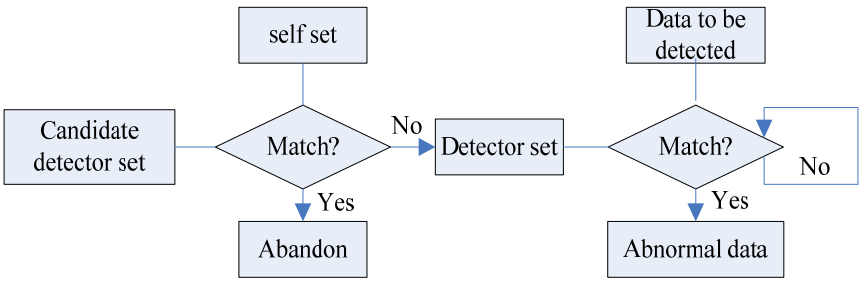

Fig. 1 The training and detecting stages of detectors

Published by Atlantis Press, Paris, France.

A certain number of candidate detectors are needed for NSA. With certain detection rate, the size of candidate detector set is associations with the size of "self" set. The space complexity of this algorithm is $O(l \cdot N_s)$, and its time complexity [2] is $O(-\ln(P_f) \cdot N_S / (P_m \cdot (1-P_m)^{N_S}))$.

It is proved in the experiments that NSA is time-consuming, and amount of candidate detectors are invalid. In order to solve these problems, D haeseleer proposed a kind of more effective algorithm which takes linear time to generate detectors [2,4]. The space complexity of the algorithm is $o((l-r)^2 \cdot 2^r)$, and the time complexity is $O((l-r).N_s) + O((l-r).2^r) + O(l.N_R)$, but the disadvantage of the algorithm is that only r-continuous matching rule can be used. Though running time and system input are in linear relation with certain parameters (length of string is l and length of continuous bits is r), its running time is still associations with r and l.

Greedy algorithm proposed by D haeseleer and Forrest [2,4] improved the efficiency of linear algorithm by eliminating redundant detectors, but its disadvantages is to sacrifice the generating speed to get fully detector set, so the space and time complexity is greatly increased.

Niching strategy [3,5] took the evolution theory of human immune system to keep the effectiveness, general and diversity of antibody. Its advantage is to highlight the adaptive ability of the immune system, and enhance the population diversity and keep good individual. It can improve the search performance of the algorithm to generate detectors with niching strategy.

IDANS: An improved Algorithm of Generating Network Intrusion Detectors

Because of the disadvantages of NSA algorithm, this paper proposes IDANS algorithm which takes advantage of the strategies such as gene library and niching strategy on the evolution process from antibody to antigen to improve detection efficiency. Three aspects of the improvement are presented in the research, the first and second aspects are during the process of generating candidate detector set and the process of self-tolerance, the third one is the use of collaborative signal.

Description of improved algorithm. The improvement of candidate detector module is as follows.

NSA took completely random method to generate candidate detectors, so even candidate detectors became mature ones after the process of self-tolerance, their efficiency to detect abnormal data is still limited. So it will produce many invalid detectors. In order to resolve such problems, three kinds of methods are presented to generate candidate detectors set in the research.

Method 1, generating candidate detectors randomly: A part of candidates will be generated randomly to insure the diversity and flexibility of candidate detector set.

Method 2, encoding and making variation with the existing abnormal mode: Detector sets could have innate antibody to known viruses with this strategy. After binary encoding and crossover-mutating, some rules reference to SNORT and existing viruses were put into candidate detector set.

Method 3, the use of gene library evolution: The optimization of gene library dynamically is used for memory and mature detectors to realize the evolution from antibody to antigen, so as to improve the ability detector sets to recognize once detected viruses. The process of optimizing gene library is as Fig. 2.

The improvement of memory detector set is as follows. The number of detector sets is confirmable, when the number of memory detector set is more than the maximum, with the strategy one of the memory detector will selected to virtual gene library and then be sent to candidate detector set after crossover-mutation.

The improvement of mature detector set is as follows. When its life-span is arrival, mature detector will be converted into memory detector if its matching number Pm is more than matching threshold, and it will be deleted from mature detector set if Pm is less than setting value, otherwise, it will be sent to candidate detector set after crossover-mutation in gene Lib.

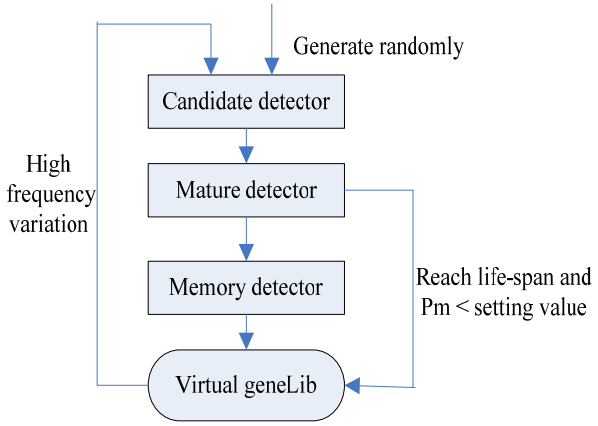

Fig. 2 The process of optimizing genelib

The improvement of redundant in generating candidate detector with NSA algorithm is as follows. Candidate detector is tolerated with existing detection set before self-tolerance, so the coverage of effective detectors will be disjoint, so as to resolve the redundant problem of detectors and improve the efficiency of generating detectors.

The use of collaborative signal is as follows. Similar to biological immune system, abnormal detection is not enough to improve immune response and it needs collaborative stimulation signal from helper cells. When detected abnormal data, memory detector will ask for triggering collaborative signal. If received stimulate collaborative signal, it will produce warning signal. And it will take the abnormal data as false alarm if suppression collaborative signal received. Otherwise, memory detector will take it as real alarm and send alarm signal if neither stimulate collaborative signal nor suppression collaborative signal were received during the setting time. With the introduction of collaborative signal, self-immune and false alarm produced by abnormal detection will be reduced, so as to improve the detection performance of the system.

Flow-process of IDANS algorithm. There are 2 stages for IDANS algorithm, which are training and detecting stage. Candidate detector set will produced and then be sent to efficient detector set after self-tolerance with the improved algorithm during training stage. And then efficient detector sets will detect network data during the process of dynamic evolution. Flow chart of algorithm is as figure 3.

Training stage is as follows.

Step1. Candidate detector set is generated according to the improved algorithm description above.

Step2. Candidate detectors produced by step1 will be deleted if they cannot be tolerated with memory and mature detector sets, otherwise going to next step.

Step3. Candidate detectors which have passed step2 will become mature detectors if they can be tolerated with self set, otherwise, they will be deleted.

Step4. Check the detector set and put the inefficient memory and mature detectors which accord with the setting condition into virtual gene library.

Step5. Repeat the steps until the setting end condition comes.

Detecting stage is as follows.

Step1. Detect the experiment data, and judge it as normal if it doesn't match the detectors. Otherwise, check whether the collaborative stimulate signal is received.

Step2. Detector sets will send alarm signal to console and record abnormal data if stimulate collaborative signal was received. Otherwise, check whether suppression collaborative signal was received.

Step3. It will be thought as false alarm if suppression collaborative signal is received. Otherwise, check whether the waiting time comes.

Step4. Alarm signal be sent to console and abnormal data be recorded if waiting time comes, otherwise, going to Step2.

Step5. Repeat the steps until the experiment data was detected all.

Published by Atlantis Press, Paris, France.

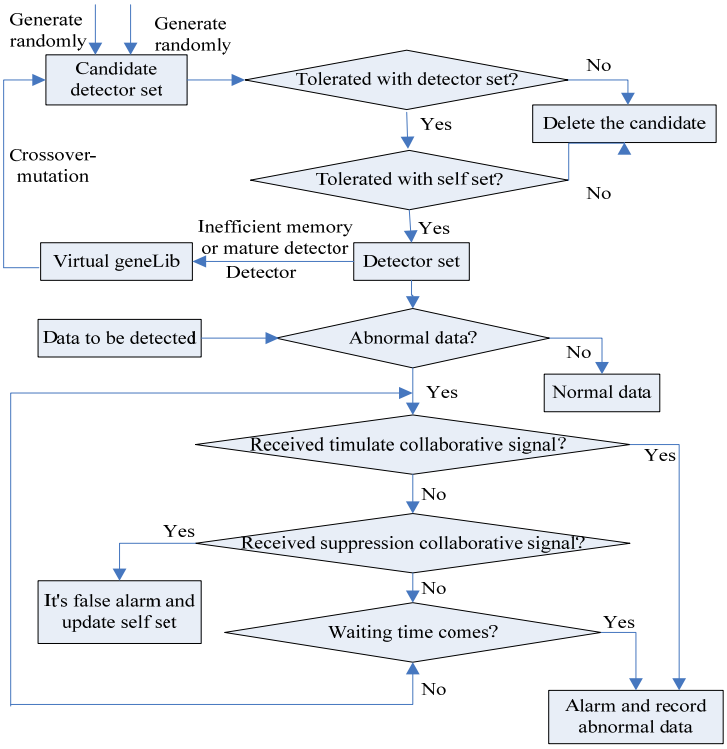

Fig. 3 The flow chart of algorithm

Analysis of IDANS. Compared with NSA, IDANS algorithm has some advantages as follows.

Encode and make variation with the existing abnormal viruses and then put them into candidate detector set, so detector sets have innate antibody to known viruses. And it can avoid producing large amount of invalid detectors.

Candidate detector is tolerated with existing detection set before self-tolerance, so as to resolve the redundant problem of detectors and improve the efficiency of generating detectors.

Optimizing of the gene library is good for detector sets to recognize once detected viruses.

The use of collaborative signal has improved the tolerance of normal changes for intrusion detection system, so as to reduce the false alarm efficiently.

Simulation Experiments

Experiment1：The data is network communicating data from a staff room' LAN in certain university. 10000 pieces of them are selected to be detection data, including 18 kinds of attack data and 8720 pieces of normal network data.

Assume that the total evolution generation is 1000, and 10000 pieces of detected data are sent to system in 10 groups. Each group includes 128 pieces of attack data and 8720 pieces of normal network data. Table 1 shows the detection result.

Table 1 Comparison of TP and FP between IDANS and NSA

Evolution generation	**NSA**		**IDANS**	
	TP %	FP %	TP %	FP %
100	97.6563	1.0321	98.4375	1.1468
200	98.0469	0.5734	98.8281	0.5734
400	98.8281	0.4587	99.2188	0.2867
800	98.2422	0.2580	99.4141	0.1147%
1000	98.4375	0.2294	99.8438	0.1147%

The result of experiment 1 shows that IDANS is better than NSA because IDANS takes collaborative signal and the strategy of updating and optimizing detector set dynamically. So,

during the evolution process, detector's ability of detecting abnormal is greatly improved. TP is nearly 99.9% and FP is nearly 0 after the evolution to 1000 generations.

Experiment2：The data chosen for the second series of experiment is available at http://www.ll.mit.edu/IST/ideval /data/1999/1999_data_index.html. The research took data of the 1st week and data of the 5th week. Data of 1st week, not including abnormal data, was used as training data. 1000 pieces of data were chosen from Data of 5th week, including abnormal data and comments, to test data.

Genes in the research is from datagram header. About 12 different kinds of virus, including 113 pieces of abnormal data were chosen to test data. Table 2 shows the detection result.

Table 2 Comparison of detection performance between IDANS and NSA

	Detected abnormal	False alarm	FP	TP
NSA	61	13	14.66%	53.98%
IDANS	86	2	2.25%	76.11%

Table2 shows the average TP and FP rates of the second series of Experiments. It shows that IDANS is still better than NSA in detection rate and false detecting rate.

Conclusion

This paper presents an improved algorithm of generating network intrusion detector (IDANS), IDANS can remove redundant detectors effectively, improve effectiveness of detector set and greatly reduce false alarm rate. Through the experiments on section 4, it is proved that IDANS is better than NSA in the detection performance.

References

[1] Stephanie Forrest, Alan S. Perelson. Self-Nonself Discrimination in a Computer[J]. IEEE, Computer Society Symposium on Research in Security and Privacy, Proceedings, Vol. 16-18 (1994), P. 202 - 212

[2] P D'haeseleer, S Forrest, Paul Helman. An immunological approach to change detection: algorithms, analysis and implications[J]. In Proc. of IEEE Symposium on Security and Privacy,. Vol. 5(1996), P. 110 - 119

[3] Ya-Jing Zhang,Chao-Zhen Hou, Fang Wang, Li-Min Su. A niching negative selective genetic algorithm for self-nonself discrimination in a computer[C]. International Conference on Machine Learning and Cybernetics, Proceedings, Vol. 11(2002) , P. 276 - 280

[4] Tao Li. Computer immunology [M]. BeiJing: publishing house of electronics industry.(2004)

[5] Zhu-Hong Zhang, Xi-Yu Huang. A kind of new immune algorithm and its application in multi-model optimizing[J]. Control theory and application,. Vol. 2(2003) , P. 17-21.

[6] http://www.ll.mit.edu/IST/ideval/data/1999/1999_data_index.html.