I.J.Computer Network and Information Security, **2010, 2, 33-39**
Published Online December 2010 in MECS (http://www.mecs-press.org/)

A new Immunity Intrusion Detection Model Based on Genetic Algorithm and Vaccine Mechanism

Jing Xiao-Pei, Wang Hou-Xiang
Information and Electric College
Naval University of Engineering
Wuhan, China
jingxiaopei@163.com

***Abstract*—After analyzing the characteristics of Immunity Intrusion Detection System, by utilizing prominent characteristics of genetic algorithm and vaccine mechanism, a new hybird immunity intrusion detection model based on genetic algorithm and vaccine mechanism was established. The modeling process is described in detail, such as feature extraction of vaccine, genetic operates to memory detectors and the improvement for detection method. Via application vaccine mechanism into intrusion detection system, the new model has the function of misuse detection and anomaly detection simultaneously. In order to improve the detection matching efficiency, we also present a novel matching algorithm RBNDM. Finally, we evaluated our model using the KDD Cup 1999 Data set. The experiments show that this model can increase the true positive rate of the IDS.**

***Index Terms*—intrusion detection; genetic algorithm; vaccine mechanism; feature extraction; genetic algorithm**

I. Introduction

Over the past several years, the principles of immunology have been adopted by the Evolutionary Computation community, resulting in the development of a new paradigm known as artificial immune systems (AIS) [1]. AIS have been applied to a wide range of problems, including computer network intrusion detection. A Classical Immunity Intrusion Detection System (IDS) tries to mimic the natural immune system by evolving a population of anomaly detectors that are capable of classifying the network traffic as normal (self) or abnormal (non-self) [2]. So the Immunity IDS has been studied by many people. In [3], the authors demonstrate how artificial immune systems use their ability to adapt to continuously changing environments. In [4], the article introduces the detector generation algorithm for artificial immune system, and divides the detector generation algorithm into three main processes, including gene library, negative selection and clone selection.

However, we know that classical Immunity IDS belongs to anomaly detection system, the immature detectors of classical Immunity IDS are generated randomly, and the time for training mature detectors is too long [5]. So the detector generation efficiency of classical Immunity IDS is lower, and the true positive rate of intrusion detection is lower too.

In this paper, a new method based on Genetic Algorithm (GA) and vaccine mechanism has been used for intrusion detection. By using this new method, we can improve the generating efficiency of candidate detectors and reduce the false positive rate of classical Immunity IDS. The rest of the paper is organized as follows: In the next section we provide a general overview on GA and vaccine theory. In Section III we propose an immunity intrusion detection model based on GA and vaccine mechanism. Section IV introduces the key technology and algorithm of the new model. A new R-continuous bits matching rule based on BNDM is described in Section V. In Section VI we do the experiment and make a discussion of the results. The paper conclusion is in Section VII.

II. Overview of Genetic Algorithm and Vaccine Mechanism

A. overview of Genetic Algorithm

GA is an efficient search method based on principles of natural selection and population genetics [6]. It is being successfully applied to problems in business, engineering and science. GA uses randomized operators operating over a population of candidate to generate new points in the search space.

GA uses three operators [7]: *selection*, *crossover* and *mutation*. The selection operator identifies the fittest individuals of the current population to serve as parents of the next generation. The primary exploration mechanism for GA is crossover. This operator randomly chooses a pair of individuals among those previous selected to breed and exchanges substrings between them. The mutation operator's main function is to restore diversity that may be lost from the repeated application of selection and crossover. This operator takes one string from the population and randomly alters some value within it. As a new algorithm of global optimize search, GA has application in many regions.

The main research of this article is the application of GA to the process of detector generating. Via implement genetic operator to memory detectors, the quality of the detector generating will be enhanced, and the efficient of detection will be improved. This method can fetch up the limitation of the random detector generating algorithm.

B. overview of vaccine mechanism

As we all know, some diseases humans cannot resist depends on peoples self immunity system, such as variola and infantile paralysis. Then people need vaccination to strengthen immunity system and prevent disease. The vaccine can supply numbers of innocuous antigens for immunity system, when the immunity system discovered antigen similar with these antigens, the immunity system will be able to produce an immune response for this kinds antigens, and create antibody to resist antigens.

Immunity system has an special memory function, if the system expose to some bacteria or virus specifically, the system can has immunity ability to these bacteria or virus last many years even lifetime [8]. So when immunity system encounters similar kinds of diseases, the system can produce an immune response quickly.

The research of this article regards known intrusion types as known diseases, and then makes feature extraction for known intrusion types to generate vaccines, and application vaccines to detectors for intrusion detecting.

III. IMMUNITY INTRUSION DETECTION MODEL USING GA AND VACCINE MECHANISM

A. The Classical Immunity Detection Model

The work flow of Classical Immunity Detection Model (CIDM) is show in Fig.1.

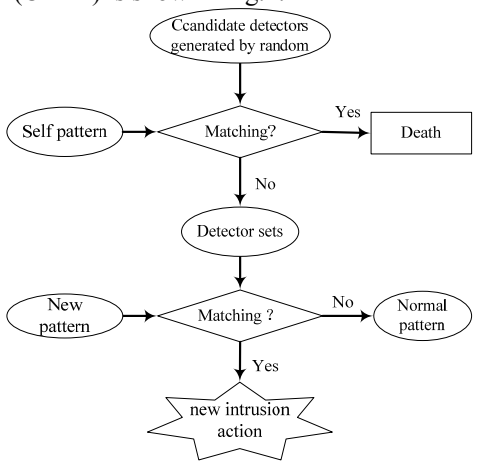

Figure 1. Flow chart of Classical Immunity Detection Model

This CIDM defines 'self' by building the normal behaviour patterns of a monitored system. Initially, it generates a number of random patterns; these patterns are exposed to self pattern for negative selection. If any randomly generated pattern matches a self pattern, this pattern fails to become a detector and thus it is removed. Otherwise, it becomes a 'detector' pattern and monitors subsequent profiled patterns of the system. During the intrusion detecting stage, if a 'detector' pattern matches any newly profiled pattern, it is then considered that new intrusion action must have occurred in the monitored system [9]. This negative selection algorithm enables CIDM to detect novel attacks. And can adapt to dynamically changing environments.

However, this model also has many disadvantages. One disadvantage is the CIDM cannot adapt the demand of real-time detection system [10]. Because of the detectors of AIS-based IDS are generated randomly, the efficiency of detector generation is lower, and the time for training mature detectors is too long. So the CIDM cannot adapt the demand of real-time detection system. Another advantage is that the CIDM lack human-robot interactivity ability [11]. System administrator cannot change the form of detectors set when the exterior environment has changed. For example, when the system administrator known a new types of intrusion action by other approach, and our system has not discover this types of intrusion action, how to add this new intrusion action into our detection set, and make our system warning when encountered this intrusion action, however, the CIDM do not consider this situation.

B. The Architecture of IIDMGV

In order to solve the problems of CIDM, and elevate the true positive rate of CIDM, we establish a novel Immunity Intrusion Detection Model using GA and Vaccine mechanism (IIDMGV). This model regards known intrusion types as known diseases, then make feature extraction for known intrusion types to generate vaccines, and application vaccines to detectors for intrusion detecting [12]. So the IIDMGV can have the function of misuse detection and anomaly detection simultaneously. The improved function of IIDMGV is described as follows:

1) The generating method for detectors generating is improved. We use two ways to generate detector set. One way is the classic method that generating immature detectors set randomly, another way is application GA in memory detectors to generate immature detectors set, this method repairs the limitation of the random detector generating algorithm, make the IDS has a strong detection ability of known viruses and attacks, shorten the training time of IIDMGV. We defined detectors sets generated by these two ways as immune detectors set.

2) Introduce vaccine detectors set. The process of vaccine detectors set generating is shown in Fig.2.

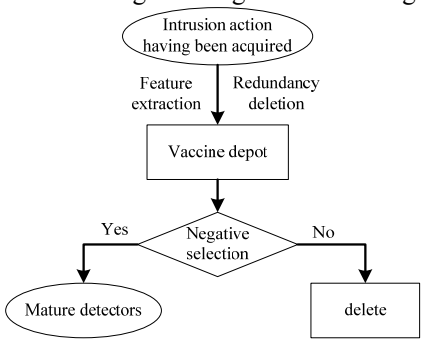

Figure 2. The process of vaccine extraction

After feature extraction and redundancy deletion for known attack actions, the immature vaccine detectors is generated, then via negative selection, Immature vaccine detectors that have survived the negative selection become mature vaccine detectors, thus, these detectors can use for intrusion detection. We defined

detectors set generated by vaccine as vaccine detectors set.

The IIDMGV is no longer an isolation system, when the Vaccine mechanism is join in it. This model can receive new detectors from other IDS to improve itself detection efficiency. This model also combines the function of misuse detection and anomaly detection. Fig.3 shows the architecture of IIDMGV.

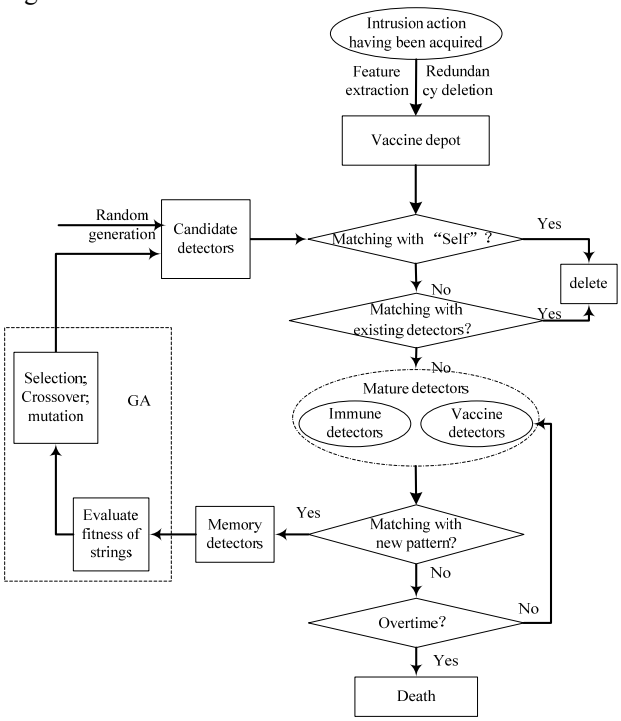

Figure 3. The architecture of IIDMGV

IV. Description of Key Technology and Algorithm

A. Basic Definition

Let U denote the finite set of binary character string which length is m, it has two subsets: self set and nonself set. Where $S \cup N=U$, $S \cap N=\Phi$. The process of detecting intrusion actions like classify binary character string I ($I \in U$), using method to distinguish I is belong to self set or nonself set. We defined detector set as P: $P= \{p_1, p_2 ... p_i\}$, $p_i \in \{0, m\}$, $i \le m$; set matching function f as: $f(I; p)$; where $p \in P$; set r as matching threshold. For anyone character string T ($T \in U$), we use (1) to justify which set that character string I belongs to.

$$match(f, \mathrm{r}, I, P) = \begin{cases} \text{nonself} & \mathrm{f}(I, p) \ge r \\ \text{self} & else \end{cases} \qquad (1)$$

B. Vaccine Detector Set Generating

In order to generate vaccine detector set, we need to do feature extraction and redundancy deletion for known attack actions. The feature extraction operator is defined as $f_e \in F$; $F= \{0, 1, *\}^l (l \in N, l>0)$, and the length is l. We set fe^k denotes the k-th bin of f_e. Let $a_1, a_2 ... a_s$ denote the known attack actions set, then the feature extraction equation is:

$$fe^k = \begin{cases} 1, & \frac{1}{s}\sum_{i=1}^{s} a_i^k \ge \alpha \\ 0, & \frac{1}{s}\sum_{i=1}^{s} a_i^k \le \beta \\ *, & \beta < \frac{1}{s}\sum_{i=1}^{s} a_i^k < \alpha \end{cases} \qquad (2)$$

Where $\alpha=0.8$, $\beta=0.2$ [13]. By observation, it can be saw that the feature extraction is to find out the common feature bin. The bin which not equals $*$ is the common choiceness feature bin. After feature extraction, we do vaccination to generate new vaccine, Let V_a denote vaccine set, and defined known intrusion set as: a_1, $a_2 ... a_s$, the vaccine generating equation is:

$$Va^k = fe^k \Theta a^k = \begin{cases} fe^k, & fe^k = 1 \quad or \quad 0 \\ a^k, & fe^k = * \end{cases} \qquad (3)$$

From (3), we can describe the characters of vaccine detectors set as follows:

1) Via feature extraction for intrusion actions, reduces the redundancy of vaccine set, by this way the detection time will be shorten, and the detection efficiency will be improved largely.

2) The vaccine reserve the common choiceness feature bin of known intrusion set, make sure that these types of attack events will be discovered quickly.

C. Improved GA in Candidate Detectors Generating

1) Coding rule

The evolution mechanism of GA is based on data coding, the search ability and population diversity of GA will be different because of data coding rule. We often use two kinds of coding rule [6]: *binary coding* and *floating point numbers coding*. In this article we choose binary coding for its stronger search ability.

As a controls parameter of GA, the size of original population (N) can affect the arithmetic astringency [7]. If the population size is too large, the time for arithmetic convergence will be longer; but if the population size is too small, it will be difficult to find the best result. In this article, the size of original population (N) is between 30 and 100.

2) Fitness function

In the process of evolution research, the fitness of an individual is estimated only by fitness function. The individual which has better fitness has high probability to become parents. The detection efficiency of the detector is also determined by fitness function. So a fitness function must be devised for each problem to be solved. In this article, individual fitness can confirm by true positive rate (TPR) and false positive rate (FPR).

true positive rate = the number of intrusion event which be detected / the number of total intrusion event;

false positive rate = the number of normal event which be detected / the number of total normal event。

Set D as the detector set, x_i is an individual of the set, $x_i \in D$. Define the individual detection value $f(x_i)$:

$$f(x_i) = TPR - FPR \qquad (4)$$

Where $f(x_i)$ is the individual detection value, the individual which has higher true positive rate can has a higher detection value. The range of detection value is [-1, 1] ($f(x_i) \in [-1,1]$). Considering that detector set D has lots of individual, the individual which can become into parents its detection value must higher than the average value of set. Therefore, we computed the average value of set firstly, then use this number subtracted from individual detection value, this formula we finally have is fitness function of the article. Because of the value of fitness function must be non-negative, so we add a number to the function, make sure the value of fitness function is non-negative.

$$F(x_i) = f(x_i) - \sum_{i=1}^{n} \frac{f(x_i)}{n} + 2 \qquad (5)$$

Where $F(x_i)$ is the fitness function. The value of the function is higher shows that the number of vulnerabilities the individual covers is more. So this individual can be selected as parents.

3) Improved Genetic Operator

The selection operator identifies the fittest individuals of the current population to serve as parents of the next generation. The selection mechanism can take many forms, but it always ensures that the best individuals have a higher probability to be selected to reproduce the new generation. In this article, we select the individuals which have higher fitness.

For any $x_i \in D$, the fitness of x_i is $F(x_i)$, so the selection probability of x_i is:

$$P(x_i) = \frac{F(x_i)}{\sum_{i=1}^{n} F(x_i)} \qquad (6)$$

Where $\sum_{i=1}^{n} F(x_i)$ is the sum of the fitness function value in the dataset D. From (6), we can see that the value of $P(x_i)$ is higher shows that the individual has more probability to be selected as parents.

After selecting the originality cluster, we need setting crossover probability Pc and mutation probability Pm, then using genetic operate to generate new CDs.

The primary exploration mechanism for GA is crossover operator. This operator randomly chooses a pairs individuals among those previously selected to breed and exchange substrings between them. The exchange occurs around randomly selected crossing points. Binary crossover has three main forms: *single-point crossover*; *multiple-point crossover* and *uniform crossover*. In this article we use three-point crossover. The main function of mutation operator is to restore diversity that may lose from the repeated application of selection and crossover. This operator takes one string from the population and randomly alters some value within it. Following nature's example, the probability of applying the mutation operator is very low compared to the probability of applying the crossover operator.

D. Improved Intrusion detecting method

The detectors set of IIDMGV have three subsets: immune detectors sets, vaccine detectors sets and memory detectors sets. The testing data which is matched by immune detectors cannot be considered as intrusion action, unless it also received coordinated response. However, the testing data which is matched by vaccine detectors or memory detectors can be considered as intrusion action, and do not need coordinated response. So the detecting steps of IIDMGV can be described in Fig.4.

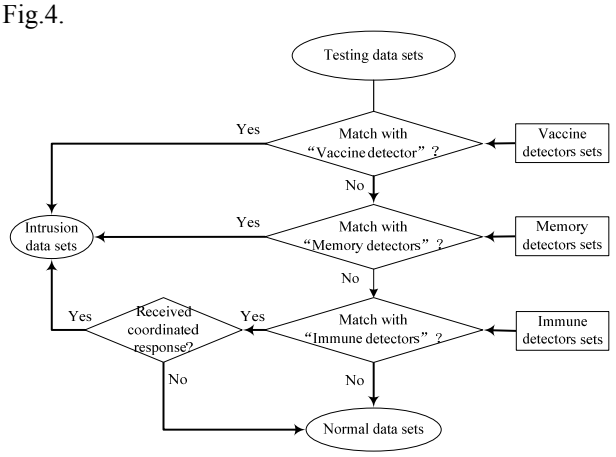

Figure 4. The detecting steps of IIDMGV

Fig.4 shows that when the model reached steady state, the form of vaccine detectors set and memory detectors set will be perfect, and most intrusion data can be detected by them without coordinated response, so the detection efficiency of IIDMGV is improved and hte time consuming of detection is reduced.

V. PATTERN MATCHING ALGORITHMS

String matching is a very important part in intrusion detecting based on Artificial Immune System, r-continuous bits matching rule is the primary matching rule in present study. However, the time consuming of the algorithm with r-continuous bits matching rule has an exponent rise with the increase of r. with the increase of r, the system will be affected greatly and the performance of system will becomes lower. In order to resolve this problem, we present a new R-continuous bits matching rule based on Backward Nondeterministic DAWG Matching algorithm (RBNDM).

A. Backward Nondeterministic DAWG Matching (BNDM)

We use the following notations. Let a pattern $P = p_1p_2 \ldots p_m$ and a text $T = t_1t_2 \ldots t_n$ be two strings over a finite alphabet Σ. The task of string matching is to find all occurrences of P in T .

The BNDM algorithm uses a table B which, for each character a, stores a bit mask expressing its locations in the pattern. The search state is kept in a word $d=d_{m-1}\ldots d_0$, The bit d_i at iteration k is set if an only if $p[m-i \ldots m-1-i+k]=t[j+m-k \ldots j+m-1]$. At iteration 0, d is set to 1^{m-1}.

The formula to update d follows $d' = (d \,\&\, B[t_j]) << 1$. There is a match if and only if, after iteration m, it holds $d_{m-1}=1$. Whenever $d_{m-1}=1$, the algorithm has matched a prefix of the pattern in the current window position j. The longuest prefix matched gives the shift to the next position. The basic version of BNDM works for patterns which are not longer than the machine word size [15, 16].

B. RBNDM

We have known that the BNDM has better matching performance, and time consuming is not large. But the theory basic of BNDM is that the all bits of pattern must match with the text continuously, however, r-continuous bits matching rule only needs r-continuous bits of pattern matching with the text. So when use BNDM for r-continuous bits matching, we need do some improvement to BNDM. The pseudocode of RBNDM is shown as below.

```
RBNDM(P = p1p2 . . . p, T = t1t2 . . . tn)
preprocessing
    for a ∈ Σ do B[a] ← 0 end of for
    for j ∈ 1...m do
    B[pj ] ← B[pj ] | (0^(j-1) 1 0^(m-j)) end of for
Searching
    pos ← 0
    while pos ≤ n − m do
        j ← r; D ←1^m
        while D ≠ 0^m do
            D ← D & B[t_(pos+j) ]
            j ← j − 1
            if j= 0 then
                report occurrence at pos + 1
            end of if
            D ← D <<1
        End of while
        pos ←pos + j+1
    end of while
```

The RBNDM algorithms also use the bit mask table B and state vector D of BNDM, we set matching threshold is r, for each pattern P, the process of string matching can be described as follow:

We first align the right ends of the pattern and the position r of the text. Then compute D form right to left, this specific work is called an attempt. In each attempt, if $D \neq 0^m$ and $j=0$, then report an occurrence at $pos + 1$, in other cases, shift the window to the right and the shift distance is j+1. In RBNDM algorithm, the maximal shift distance of RBNDM is r.

VI. Experiments and Analysis

A. Data preprocess for experiment dataset

The test dataset we selected is from KDD Cup 1999 Data, the dataset was obtained from MIT Lincoln laboratory 1998 DARPA Intrusion Detection Evaluation Dataset. This database contains a standard set of data to be audited, which includes a wide variety of intrusions simulated in a military network environment [17].

The data from KDD Cup 1999 can be partitioned into four kinds of attack: (1) Probe attack, e.g. Ipsweep, Mscan, Nmap etc; (2) DoS (Denial of Service) attack, e.g. Apaeh2, Ping of Dath, Smurf, Teardrop etc; (3) U2R (User to Root) attack, e.g. Eject, Ffbconfig, Perl etc; (4) R2L (User to Local) attack, e.g. Dictionary, Ftp-write, Imap etc.

The records from KDD Cup 1999 has 41 features, but only 31 features can representation the variety of various records, such as: duration—the connection lasting time; protocol-type, e.g. tcp, udp etc; service—target network service, e.g. http, telnet etc[18]. Due to the data features of KDD is too large, and some features are not very important for intrusion detecting. Therefore, we can select the features which the value will be changed distinctly when being attacked as intrusion info. After analysis the 31 data features, and based on the work by [13], we select 12 features to use for intrusion detection, these features are fall into three categories, and the features we selected are shown in Table I.

TABLE I.
THE FEATURES SELECTED FOR INTRUSION DETECTION

category	feature name	type
Basic features of individual TCP connections	duration	continuous
	protocol_type	discrete
	service	discrete
	flag	discrete
Content features within a connection suggested by domain knowledge	hot	continuous
	num_failed_logins	continuous
	root_shell	discrete
	num_access_files	continuous
Traffic features computed using a two-second time window	count	continuous
	srv_count	continuous
	diff_srv_rate	continuous
	srv_diff_host_rate	continuous

However, because of these features have different data types, such as: continuous and discrete. So we need further process for these feature data. For discrete type features, we just transform it into binary code. For continuous type features, we need normalized process firstly, the normalized function we use is

$$f(x)=\begin{cases}(x-x_{\min})/(x_{\max}-x_{\min}) & x_{\max}\neq x_{\min}\\ 0 & x_{\max}=x_{\min}\end{cases} \quad (7)$$

Where $x \in [x_{min}, x_{max}]$, from (7), we can see that $f(x) \in [0, 1]$. We keep two decimal places for $f(x)$, and transform it into binary code which length is 8. Finally, each record we selected will be transform into binary code which length is 72.

B. Experiment and Analysis

We set the size of self population $N=4000$, testing population $T=3000$, vaccine population $V_a=200$, matching threshold $r=10$, probability of crossover

P_c=*0.8*, probability of mutation P_m=*0.1*. The alternate time for testing is *5000*. After setting the experimentation parameter, make use of WEKA software to do data experimentation. In the article, In order to show the better performance of IIDMGV, we compared the performance between CIDM and IIDMGV, and the experimentation parameters of two models are setting the same. The results for these two models are show in Table II.

TABLE II.
RESULTS FOR TWO KINDS OF DETECTION MODELS

Intrusion Types	CIDM		IIDMGV	
	TPR	***FPR***	***TPR***	***FPR***
Probe	93.32%	5.35%	98.51%	3.21%
DoS	94.52%	6.67%	99.55%	1.02%
U2R	92.67%	4.63%	97.89%	1.53%
R2L	93.58%	6.58%	97.12%	2.12%

As seen from Table, the performance of IIDMGV is better than CIDM. This is because two mainly factors: (a). Application GA operates to CDs generating process. Via this way the quality of the detector generating is enhanced, and the detection efficient is also improved; (b). Introduce Vaccine mechanism into intrusion detecting. This method make the form of IIDMGV detectors much more perfect, and because of this method the model have the function of misuse detection and anomaly detection simultaneously. Because of introduces these two innovative methods, the IIDMGV has better self-adapt ability, and make the model has strong detection ability of viruses and attacks which has been discovered

From the Table we also see that the model has better detection effect for DoS, this is because the proportion of DoS detectors in vaccine is larger. In this way we can improve the detection effect for some kind's intrusion types purposefully, by control the proportion of these intrusion types in vaccine. Via application Vaccine mechanism to the model, the human-robot interactivity of model is strengthened obviously.

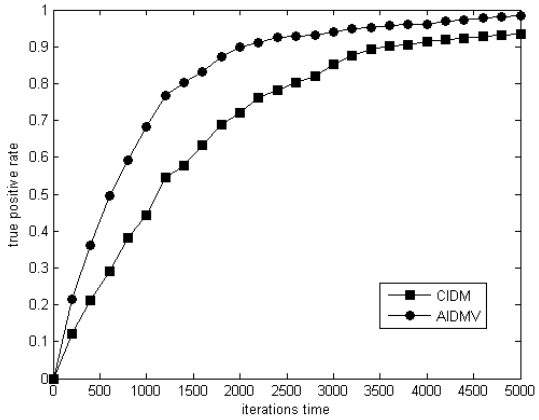

Figure 5. The relationship between the iterations time and the TPR

In order to know the relationship between the number of iterations time and the performance of IIDMGV, we set different number of alternate time to see the change of IIDMGV performance. We displayed the result in Fig.5. The result shows that when the number of iterations time increase, the detection performance is better relatively. And along with the iterations time increasing, the IIDMGV reach the stabilization state faster than CIDM, this means the detection algorithm of IIDMGV has better convergence.

VII. CONCLUSION

Because of the diversification of intrusion actions, a traditional immunity intrusion system, based on anomaly detection theory, cannot detect attacks with higher efficiency. A new developing direction of intrusion detection technology—hybrid intrusion detection technology has been researched by more and more people.

The article puts forward a novel immunity intrusion detection model using genetic algorithm and vaccine mechanism. Via application GA and a vaccine mechanism into intrusion detection system, change the generating mechanism of mature detectors and improve the composing of detectors population, the IIDMGV have the function of misuse detection and anomaly detection simultaneously. Finally, we implement experiment for the new model, and the experiments show this model is efficient, and can be used in practice to monitor the computer system in real time. In future research, how to optimize the detector set size and reduce the false positive rate will be paid more attention.

REFERENCES

[1] Hofmeyr, S, Forrest, S. "Immunity by Design: An Artificial Immune System," Proceedings of the 1999 Genetic and Evolutionary Computation Conference, 1999, pp. 1289-1296.

[2] Jiao Li-cheng, Du Hai-feng, et al. "Immune Optimization Computing, Learning and Identification," Bei Jing: Publishing House of Science, 2006, pp. 89-96.

[3] Jungwon Kim, Peter J. Bentley. "Towards an Artificial Immune System for Network Intrusion Detection: An Investigation of Dynamic Clonal Selection," wcci, 2002, vol. 2, pp.1015-1020.

[4] Jinyin Chen, Dongyong Yang, et al. "A Study of Detector Generation Algorithms Based on Artificial Immune in Intrusion Detection System," WSEAS TRANSACTIONS on BIOLOGY and BIOMEDICINE, 2007, vol. 4 (3), pp.29-35.

[5] Hunt J E, Cooke D E. "Learning Using an Artificial Immune System," Journal of network and computer applications, 1996, vol. 19, pp. 189-212.

[6] Cantu-Paz E. "Feature subset selection, class separability, and genetic algorithms," Proceedings of the Genetic and Evolutionary Computation Conf, 2004, pp. 959-970.

[7] Kop M, Liu Xiu-fen. "Texture Detection by Genetic Programming," Proceedings of Congress on Evolutionary Computation, 2001, pp. 867-872.

[8] Hou Hai-yu, Gerry Dozier. "Immunity-Based Intrusion Detection System Design, Vulnerability Analysis, and GENERTIA's Genetic Arms Race," ACM Symposium on Applied Computing, 2005, pp. 952-956.

[9] Jungwon Kim, Peter J. Bentley. "An Evaluation of Negative Selection in an Artificial Immune System for Network Intrusion Detection," In GECC0-2001 Proceedings, 2001, pp. 1330–1337.

[10] Hofneyr S A, Forrest S. "Architecture for an artificial immune system," Evolutionary Computation, 2000, vol. 8(4), pp. 443-473.

[11] Chen You, Cheng Xue-Qi, Li Yang, et al. "Lightweight Intrusion Detection System Based on Feature Selection," Journal of Software, 2007, vol. 18(7), pp. 1639-1651.

[12] Jing Xiao-pei, Wang Hou-xiang, Han Ruo-fei, et al. "Improved Genetic Algorithm in Intrusion Detection Model Based on Artificial Immune Theory," Proceedings of the CNMT 2009, 2009, pp. 659-662.

[13] Yan Xuan-hui. "An Artificial Immune-Based Intrusion Detection Model Using Vaccination Strategy," Acta Electronica Sinica, 2009, vol. 37(4), pp. 780-785.

[14] Yu Y, Huang H. "An ensemble approach to intrusion detection based on improved multi-objective genetic Algorithm," Journal of Software, 2007, vol. 18(6), pp. 1369-1378.

[15] Gonzalo Navarro, Mathieu Raffinot. "Flexible Pattern Matching in String," Bei Jing: Publishing House of Electronics Industry, 2007, pp. 26-29.

[16] Branislav Durian, Jan Holub, Hannu Peltola, Jorma Tarhio. "Tuning BNDM with *q*-Grams," In proc. of Tength Workshops on Algorithm Engineering and Experiments, 2009, pp. 29-37.

[17] KDD cup 1999 data. http://kdd.ics.uci.edu/databases/ kddcup99/ kddcup99.html.

[18] Wang Jie-song, Zhang Xiao-fei. "The Analysis and Preprocess for Network Intrusion Detection data KDDCup99," Science & Technology Information, 2008. 15, pp. 407-408.

