The Clonal Expansion and Memory Strategy Applied to Network Detection

Yuling Tian
College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan Shanxi 030024, China
Email: tianyuling@tyut.edu.cn

Abstract—The dynamic tendency of network environment determines that system can achieve an accurate fault diagnosis only by self-learning. Inspired by characters of artificial immune and adaptability of dynamic clonal selection algorithm for dynamic environment, an immune algorithm applied to network fault diagnosis was proposed based on the detector population quality and the memory characteristics. The clonal expansion strategy was designed to improve the quality of mature detector populations and the classification memory strategy can achieve dynamic updated memory detector population through evaluating the effectiveness of the memory detectors. The experimental results show that the network fault diagnosis based on immune theory can achieve successive learning to accommodate the emerging new situations, and improve the accuracy rate and efficiency of detecting known and unknown faults.

Index Terms—immune algorithm, clonal expansion, memory classification, fault detection

I. INTRODUCTION

The artificial immune system is a kind of theory and technique which is inspired by biological principles, and widely used in network security, fault diagnosis, machine learning, pattern recognition and data mining due to its excellent characteristics such as self-adapting, self-organization and immune recognition [1]-[3]. In the field of network security, existing researches, focusing on the network dynamic environment, relevant algorithms and models based on immune mechanisms have been designed [4]-[6]. For example, Hofmeyr conducted beneficial extension on artificial immune system, in 1999, which makes the target system learn on the new condition appeared in network environments and generate new detectors for abnormal detections [7]. Inspired by research achievement of Hofmeyr, Kim and Bentley have proposed the dynamic clonal selection algorithm (DynamiCS) [8], [9], which can constantly detect changing network intrusion activities by learning self and forecasting non-self dynamically. The algorithm proposed in literature [10] extends DynamiCS by eliminating invalid memory detectors, and decreases the false positive (FP) error rates caused by memory detectors.

Due to the mutual restraints of the traditional dynamics immature and mature detectors’ population quantity, the detection rate cannot achieve satisfactory results. Inspired by the mechanism of clonal selection and using the dynamics operating mechanism for reference, an immune algorithm based on the cloning expansion strategy is proposed. This algorithm makes a detail design for the cloning expansion strategy of mature detectors, the generation of initial detectors, the self-tolerance and the supplementary conditions of immature detectors. It breaks the bind effects of immature and mature detectors’ population quantity, and designs a method for mature detectors’ efficiency assessment by adjusting the supplementary conditions of immature detectors. Since the effectiveness of detectors determines the cloning scale, the mature detector population which has higher effectiveness is taken into cloning expansion strategy. It improves the quantity and quality of memory detectors by increasing the quantity of mature detectors appropriately. Finally, the algorithm’s detection rate is improved. The false alarm rate is suppressed, and the diversity of detectors population is enhanced with better adaptability.

II. DYNAMIC CLONAL SELECTION ALGORITHM

In network-based intrusion detection system, it needs to monitor a real environment which produces new network traffic continuously in real time. The normal behaviour of network traffic on one day, which is considered as self antigens, can be different from normal behaviour of network traffic on another day. Therefore, the antigens faced by the artificial immune system will be different every day, the artificial immune system needs to be extended. Describes a dynamic clonal selection algorithm that has the two properties:

- Firstly to learn normal behaviour by undergoing only a small subset of self antigens at one time.
- Secondly its detectors should be replaced whenever previously observed normal behaviour no longer represent current normal behaviour.

The dynamic clonal selection algorithm was achieved via coordinated dynamics of three different detector populations: immature, mature, and memory detector populations.

DynamiCS starts by seeding initial immature detectors with random genotypes, and then employs negative selection by comparing immature detectors to the given antigen set. As the result, immature detectors that bind to any antigens are deleted from the immature detector population and new immature detectors are generated.
until the number of immature detectors becomes the maximum size of the non-memory detector population. The same processes continue for the tolerance period (T) number of generations. When the total number of generations reaches T, those immature detectors whose age reaches T (born at generation 1), become mature detectors.

At generation \( T + 1 \), a new antigen set is presented to the mature detectors to be monitored. Whenever a mature detector matches an antigen, the match count of a mature detector increases by one. After all the given antigens have been compared to all the existing mature detectors, the system checks: finally, intervenes the evolution and searching direction of detectors.

- Whether the match counts of mature detectors are larger than a pre-defined activation threshold (\( A \)).
- If there is a mature detector with a match count that is larger than \( A \), this mature detector becomes a memory detector only if it indeed detects an intrusion. When a human security officer acknowledges that this detector detects any intrusion signature (co-stimulation), the detector activates and eventually becomes a memory detector.
- In addition, if the ages of mature detectors meet \( L \), those mature detectors are deleted from the mature detector population.

At generation \( T + 2 \), when memory detectors match any antigen, confirmation is sought immediately from a human security officer. In this case, if the detected antigen patterns are confirmed as intrusion signatures, the detected antigen patterns are instantly deleted from the antigen set.

III. THE EXPANSION AND MEMORY STRATEGY OF DETECTORS

A. Related Definitions

The detectors’ detection ability changes greatly when they are in evolution which is the improvement or the deprivation of detection ability. The efficiency assessment provides an access to forecasting detection ability so as to single out detectors with higher quality and finally involves the evolution by searching direction of detectors.

- **Detector population**
  
  Set \( Ab \) as detector population. \( Ab = \{ \omega | \omega \leq r, age, life, match, r \in Ag, age, life, match \in N \} \) where \( r \) is the detector and \( age, life, match \), represents the age, survival time and the times that detector matches with antigens respectively. \( N \) is the natural number set and \( Ag \) is the antigen set containing self set \( self \) and non-self set \( non_self \) which meet the condition that \( self \cup non_self = Ag, self \cap non_self = \emptyset \).

- **Memory detector population**
  
  Set \( R \) as memory detector population. \( R = \{ x | x \in Ab, x.match \geq A \} \), where \( A \) is the activation threshold and \( x = \{ x_1, x_2, ..., x_n \} \) is \( n \)-tuple.

- **Mature detector population**
  
  Set \( M \) as mature detector population. \( M = \{ x | x \in Ab, x.life < L \} \), where \( L \) is the life-cycle of mature detectors and \( x = \{ x_1, x_2, ..., x_n \} \) is \( n \)-tuple.

- **Immature detector population**
  
  Set \( I \) as immature detector population. \( I = \{ x | x \in Ab, x.age < T \} \), where \( T \) is the tolerance time and \( x = \{ x_1, x_2, ..., x_n \} \) is \( n \)-tuple.

- **Effectiveness evaluation**

  The efficiency of detectors means the detectors’ ability of detecting abnormal behaviors. In unit time, the more antigens detectors matched with, the more efficient it will be. The effectiveness evaluation provides a method to calculate the detection ability of the detector. Given a detector as \( x \), then its efficiency \( P(x) \) will be

  \[
  P(x) = \frac{x.match}{x.age} \tag{1}
  \]

  where \( x.age \) represents the age of the detector \( x \) and \( x.match \) represents the number that \( x \) matches with antigens. A higher value of \( P(x) \) shows a greater detecting ability of \( x \). If there exists \( P(x_1) = P(x_2) \) which is calculated by the formula (1) and \( x_1.age < x_2.age \), then set \( P(x_1) < P(x_2) \).

B. Clonal Expansion Strategy

The clonal expansion of mature detectors based on the result of efficiency assessment is a value added process of high efficiency mature detectors and also a learning or optimization process. Its essence is to generate a mutation detector population and amplify the search space, according to the result of efficiency assessment, near mature detectors in the next mutation generation. After mature detectors experienced the clone and mutation, the detecting ability has promoted, and the quantity of high quality detectors has increased, so the probability of becoming memory detectors has increased. The clonal expansion strategy has mainly four steps including choosing detectors with high efficiency, conducting cloning process on the efficient detectors, conducting mutation process on the cloning population and eliminating the redundant mature detectors.

Specific description as follows:

Step 1 Choosing factor.

The Choosing factor operation is for the generating of the clonal expansion of mature detectors with high quality. Each detector in mature detector population has conducted effectiveness calculation by formula (1), let the detector \( x_i(i = 1, 2, ..., n) \) array be in descending order according to the effectiveness \( P(x_i)(i = 1, 2, ..., n) \), and then choose the mature detector subset \( M' \) composed of the preceding \( t \) detectors.

Step 2 Clone factor.

Cloning factor could make an immune system’s response more rapid and effective. In this process, those antibodies with high quality will retain and generate more descendants. The cloning factor for mature detector subset
will lead to the increment of mature detector \( M'' \). Let the detector set after cloning operation be \( M'' \).

\[
M'' = C(M') = [C(x_1), C(x_2), \ldots, C(x_t)]^T
\]

(2)

where \( C(x_i) = \lambda_i \times x_i \), \( \lambda_i \) is \( \zeta_i \), dimensional vector and \( \zeta_i \) is the individual’s cloning scale. Let the detectors be in \( M'' \) array in descending order according to the effectiveness. And let \( i \) in the ranked first detector to be 1, with this analogize.

\[
\zeta_i = \text{round}(\beta \times N/c_i)
\]

(3)

where \( \beta \) is a propagate gene and \( N \) is total quantity of the existing antibodies. Then the total quantity of cloned detectors will be

\[
N_c = \sum_{i=1}^{t} \text{round}(\beta \times N/c_i)
\]

(4)

Step 3 Mutation.

Because immune system experienced mutation has the abilities of expansion, dynamic adaptation and self-learning which is more rapid and effective in the process of antigen recognition, the antibodies after cloning operation will experience mutation. Gaussian mutation is learning which is more rapid and effective in the process of expansion, dynamic adaptation and self-learning.

Conduct the below operations on \( x_i(j), i = 1, 2, \ldots, n \) which is the component of detector \( x_i(i = 1, 2, \ldots, n) \).

\[
x_i'(j) = x_i(j) + \eta_j^i N(0, 1)
\]

(5)

\[
\eta_j^i = \eta_j^i \cdot \exp(\tau \cdot N(0, 1) + \tau \cdot N_i(0, 1))
\]

(6)

where \( x_i(j), x_i'(j), \eta_j^i, \eta_j^i \) represents the \( j \)th component of vector \( a_i, a_i', \eta_i, \eta_i' \) respectively and \( N_i(0, 1) \), whose mean is 0 and variance is 1, represents the normal random number generated by each corresponding \( j \). Let \( \tau \) be \((\sqrt{2N})^{-1}\) and \( i \) be \((\sqrt{2n})^{-1}\).

Step 4 Eliminate the redundancy detectors.

The quantity of mature detectors, which experienced the process of cloning and mutation, has increased greatly, and the position has been changed. The detection range of detectors may overlap, leading to the generation of redundancy detectors. So, those redundancy detectors should be eliminated. The algorithm uses the detector assessment strategy in literature [11] for reference to assess the detection range’s overlapping degree.

Defining the overlapping degree of detector \( x \) and others’ detection range as overlapping values, represented by \( W(x, x') \), that is

\[
W(x, x') = \sum_{x \neq x'} w(x, x')
\]

(7)

where \( W(x, x') \) is the overlapping values of the detection range of detector \( x \) and detector \( x' \). It shows that the overlapping values of detector \( x \) is the sum of adjacent detectors’ overlapping degree.

\[
w(x, x') = (\exp(\delta) - 1)^V
\]

(8)

where \( V \) represents the dimension of shape space,

\[
\delta = (r_{x} + r_{x'})/2r_{x}
\]

(9)

where \( r_{x} \) represents the radius of \( x \), and \( r_{x'} \) represents the radius of \( x' \). \( D \) represents the distance between the two detectors. We can perceive the overlap between the two detectors from Figure.1 directly. If the overlapping value \( W(x, x') \) of detection range of detector \( x \) is larger than threshold, then delete this detector.

C. Classified memory strategy

1) Detector classification: Immune memory is a kind of complicated defense mechanism. The research of immune memory by Yates, Callard and other scholars from different perspective shows that the amount of memory cells keeps substantially unchanged and it is a dynamic process that new memory cells emerge, meanwhile those which are lack of antigen stimulation pass away. Though many researchers promulgate the stability of the amount of memory cells, seldom does the literature value the memory cells. The effectiveness assessment mechanism proposed by this paper values the characteristics of detectors and classifies of the memory detecting population, namely \( R \), according to the assessment result, expecting increasing the detection performance.

Conducting the effectiveness evaluation for memory detector population, namely \( R \), then let the detector array in descending order according to the assessment results. Separating the memory detector population into three sub-populations according to the ratio of 7 : 2 : 1, and obviously \( R_1 \cup R_2 \cup R_3 = R \). Then the following operation would be conducted:

- The detector sub-population \( R_1 \) should be reserved as memory detectors, then conduct direct evolution strategy for detectors in \( R_1 \) with probability \( P \).
- The detectors \( x_i(i = 1, 2, \ldots, n) \) in \( R_2 \) are mapping to \( M \) so as to reevaluate it.
- Supply \( I \) with the variants of detectors in \( R_3 \) which is regarded as the virtual gene pool of immature detectors.

2) The evolution strategy of \( R_1 \): \( R_1 \) is the most effective detector set and is highly adapted to the current environment in memory detector population \( R \). Conducting the direct evolution for \( R_1 \) can eliminate the synergy stimulation in the process of transformation from \( M \) to
Conduct the operation of clone on sub-set detectors $R_1$ with probability $P$. Let the resultant detector set be $R'_1$.

$$R'_1 = C(R_1) = [C(x_1), C(x_2), \ldots C(x_3)]^T$$

where $C(x_i) = \lambda_i \times x_i$, $\lambda_i$ is $\zeta_i$ dimensional row vector and $\zeta_i$ is the detector individual’s cloning scale.

$$\zeta_i = \text{round}(\alpha \times \beta \times N)$$

where $\alpha$ is a regulatory factor, $\beta$ is a propagate factor and $N$ is the total quantity of the existing antibodies.

$$\beta = \frac{p(x)}{d(x)}$$

where $p(x)$ is the efficiency of the detector and $d(x)$ is the concentration of the detector. Then the total quantity of the cloned detectors will be

$$N_c = \sum_{i=1}^{n} \text{round}(\alpha \times \beta \times N)$$

Conduct mutation in cloned detector population. A mutation can make an immune system capable of scalability, dynamic adaptability and self-learning so as to recognize antigen rapidly. We adopt gaussian mutation which can generate the mutation near the initial detectors, has strong local search ability and can reserve the population information of initial detectors, see equation (5)(6).

3) The evolution strategy of $R_2$: The real-time variation of network environment might make the fault types in time $t_1$ and time $t_2$ quite different, while fault types in time $t_1$ and time $t_3$ is almost the same. If it matches the current environment, the memory detector sub-population $R_2$ mapping to mature detector population $M$ can be chosen as memory detectors after generations of evolution. This process significantly reduces the probability of misdiagnosis. Besides, those detectors would not have to experience the evolution from $I$ to $M$, which is beneficial to improve convergence speed.

4) The evolution strategy of $R_3$: Repeal the memory ability of detector sub-population $R_3$, in order to detect, and let it be the virtual gene pool of immature detectors. The detectors in $R_3$ reserve useful information in the process of evolution, which is contributing to find a new non-self niche. Conduct uniform mutation with higher local escape ability on them.

Single out a value $\Delta x$ in interval $[a_i, b_i]$, and add it up to, the component of detector $x_i(i = 1, 2, \ldots, n), x_i(j), i = 1, 2, \ldots, n$. That is $x'_i = x_i + \Delta x$. When the variant $x'_i$ exceeds the defined interval, then delete it.

The process of mutation can eliminate the self information of $R_3$ to a great degree. Adding variants into the immature detector population $I$ can lower the elimination rate of immature detectors and shorten the process of self tolerance.

IV. THE DESIGNATION OF IMMUNE ALGORITHM USED FOR NETWORK FAULT DIAGNOSIS

In the realistic network environment, the new behavior patterns emerge continuously. Hence, we need to make a rapid and accurate judgment to the current behavior pattern so as to ensure the alarm accuracy of the fault detection system. Memory detectors play an important role in the non-self antigen detection whose high effectiveness may improve the performance of a detection system to a great extent. In the memory detection population, some detectors with $n$th generation of behavior pattern show high efficiency while some are not at all. But some detectors which show lower efficiency in $n$th generation show high efficiency in $n + i$th generation. So, we proposes an immune algorithm based on layered memory strategy. Figure 2 shows the details.

A. The generation of initial detectors

The memory detector set is a null set in the first iteration. Take self antigens in the training antigen set as immature detectors to reduce the computation of immature detectors between mature detectors and improve
the evolution rate of immature detectors leading to less evolution time and higher convergence speed.

B. Self tolerance of immature detector

The immune tolerance is a state of systemic unresponse specifically showed by immunocompetent cells when they contact with a particular antigen, and the foreign or self antigen both can induce immune tolerance. Self tolerance is an immune tolerance that has unresponse specific to self antigen, so the immune system can distinguish self and non-self by self tolerance. The immune system would attack itself if the autoimmunity is triggered by the combination of immune cells and non-self. There are three methods to handle the cells which cannot do self tolerance: clonal deletion, clone anergy and receptor editing. T cells and B cells should experience the self tolerance during their changes from immature to mature ones. Because immature detectors do not have the ability of detecting invasion antigens, they should experience the self tolerance to turn into mature ones before detecting. The negative selection algorithm is a kind of simulation for the process of immune cells becoming mature, regarding the detector experiencing tolerance as mature detectors. This algorithm includes two stages: tolerance and detection, while the former is in charge of the generation of mature detectors. The system adopts the negative selection algorithm with mutation proposed by Castro and Timmis in 2002 to conduct the tolerance of immature detectors. Let each immature detector match with self data, if successful match, then make the immature detectors to far away from self. Figure.3 shows the details.

C. The supplement of immature detectors

There are two ways to supply for I. One of them is to supply I with detectors mutated from $R_3$. This method plays a guiding role that it contributes to accelerate the convergence rate of the algorithm. The other one is to supply I with detectors which are generated randomly, contributing to maintain the diversity of the population. Specific description as follows:

\[
\text{if}(R_3 > 0) \quad \{ \text{single out a detector } x_i (i = 1, 2, ..., n) \text{ from } R_3 \text{ randomly,}
\]
\[
\text{Conduct uniform mutation,}
\]

Supply I with the resultant detectors. }
\]
\[
\text{else} \quad \{ \text{generate a immature detector } x_i (i = 1, 2, ..., n), \text{ Supply I with detector } x_i (i = 1, 2, ..., n) \}
\]

D. The clonal expansion of mature detectors

The clonal selection theory was proposed by an Australian immunologist, Burnet, in 1958. The basic ideas may be described as follows. The invasion antigens can induce the body to generate relevant antibodies. Only those antibodies which can recognize antigens would be chosen by the immune system and then in large scale propagation, while those which cannot recognize antigens would not propagate generally. The clonal expansion strategy based on clonal selection theory would make an immune system with more rapid and efficient response. Adopting the clonal expansion strategy, the optimal individuals would be reserved and propagate in the process of clonal expansion for mature detectors.

E. The classified memory of memory detectors

A series process of regeneration and elimination of immune cells will occur during the immune system’s evolution, guaranteeing the diversity of immune cell population and acute search ability against the non-self antigen. The memory cells have long life-cycle which can rapidly eliminate the antigens that invade again, and can be activated to generate new immune cells. While partial new immune cells can transform into mature cells directly and some others become mature population by mutation and tolerance. The effectiveness assessment mechanism proposed by this paper values the characteristics of detectors and classifies the memory detecting population, namely $R$, according to the assessment result.

V. THE EXPERIMENT AND ANALYSIS

A. Experiment scheme

Test the immune algorithm which is based on clonal expansion strategy (ImmuneCE) by standard data sets, KDD-99. The data sets are divided into training and testing sets, which include 24 and 38 attack types respectively. This experiment adopts the antigen presentation method proposed in reference [12].444 normal (self) data would be selected from the data sets, and separated into 3 groups which have 196, 112, and 136 normal(self) data respectively. Experimental environments choose Matlab 2011a, windows 7, Visual C++6.0 as the simulation test tool platform.

In the same way, 444 abnormal(non-self) data would be selected, and separated into 3 groups which have 97, 69, and 73 abnormal(non-self) data respectively. Selecting two groups of data from the above groups, in which one is self-data set and the other is non-self-data set, and then 80% data would be selected from the two groups randomly for antigen presentation.Reselecting two new groups for antigen presentation after $N$ generations’ evolution, and using the former groups circularly after $3N$
generations, which could guarantee each generation has new data. Adopting two different parameters to compare ImmuneCE with traditional DynamicCS, and setting the evolutionary generation is 2000, repetition time is 10.

B. Experimental results

Figure 4 and Figure 5 illustrate the comparison between ImmuneCE and DynamicCS about their detection rate and false alarm rate, while $T=10$, $A=15$, $L=20$ and $T=30$, $A=5$, $L=10$ respectively. TP1 and FP1 represent the detection rate and false alarm rate of ImmuneCE respectively. While TP2 and FP2 represent the detection rate and false alarm rate of DynamicCS respectively. The figures show that ImmuneCE has higher detection rate and lower false alarm rate than DynamicCS.

C. Result analysis

The figures above show that both algorithms have a fluctuation of detection rate and false alarm rate when the parameter values change, and also DynamicCS shows stronger sensitivity to parameters than ImmuneCE. In ImmuneCE, the clonal expansion of mature detectors has increased the quality of mature detectors and also the quality of memory detectors while appropriately augments the quantity of mature detectors, and ultimately enhance the detection rate.

The mature detector populations in DynamicCS are generated by the tolerance of immature detectors, so that the quality and quantity of mature detectors cannot meet the requirement, which is a bottleneck on increasing detection rate. In ImmuneCE, the clonal expansion of mature detectors has increased the quality of mature detectors and also the quality of memory detectors while appropriately raising the quantity of mature detectors, and it increases the detection rate ultimately.

Increasing the tolerance period ($T$) of immature detectors can make the immature detectors fully study on antigens, contributing to descending the FP. While for DynamicCS, it will reduce TP, because increasing the tolerance period leads to longer survival time of immature detectors, which increases the quantity of immature detectors. For the quantity of non-memory detectors in DynamicCS is determinate, so the quantity of mature detector population, a kind of candidate population which are stimulated memory detectors, is reduced, decreasing the quantity and quality of memory detector population ultimately. The quality of mature and memory detector population has a direct influence on antigen (self and non-self) matching probability. Through conducting clonal expansion strategy on detectors of high efficiency, ImmuneCE increases the quality of mature detector population and increases TP while suppressing FP.

In the two algorithms, memory detectors are generated by mature detectors, which have achieved activation threshold, processing synergy stimulation. Lower activation threshold ($A$) can increase the quantity of mature detector populations which have achieved it, so that there will be more memory detector population, leading to the high antigen matching probability of memory detector population, that is to increase TP. And Lower activation threshold ($A$) can depress the quality of memory detectors, making FP increase.

Apparently, increasing the life-cycle ($L$) of mature detectors is helpful to enhance the quantity of mature detector population, and make mature detector have more opportunity to match antigen at the same time, which improves the probability of transforming mature detectors into memory detectors, so that the quantity of memory detector increases and thus enhance TP. ImmuneCE changes the supplementary condition of immature detectors and releases the restrictive relationship between mature and immature detectors in quantity, thus increases the detection rate while suppresses the false alarm rate.

VI. Conclusion

The clonal expansion strategy and classified memory strategy are proposed which applied to the design of network fault diagnosis model. The two have achieved the learning and evolution of mature detector pool and
memory pool. Applying the data, which are obtained by processing collected network operation information, to the performance verification of the model, and the performance verification results show that network fault diagnosis model based on immunization theory can achieve effective detection and learn about known and unknown faults. In the later research work, integrating the neural network, immunization theory, expert system, etc, and utilizing each others' advantages, a new method of fault management will be achieved.

REFERENCES


Yuling Tian (1963-), female, birthplace HuNan ZiXing, associate professor, TaiYuan university of technology, doctor, master’s supervisor, main research topics: artificial intelligence, fault detection and diagnosis.