Research on security situation assessment algorithm under virtual technology of cloud platform

Jie Liu[1,*], and *Weiping* Fan[1]
Jiangxi Radio and Television University Jiujiang Branch,Department of Teaching.Jiujiang City, Jiangxi Province,332000,China

Abstract. With the increasing attacks and destructions against cloud platforms, the security guarantee mechanism of cloud platforms has also changed from traditional passive protectionto active defense. Situation assessment is a method that can actively analyze and evaluate thecurrent security risk status of the cloud platform, and is a key part of the whole process of situation awareness. This paper aims at a large number of virtual machines deployed in cloud platforms,on the basis of analysis and extraction of virtual layer security situation assessment elements, an improved adaptive genetic simulated annealing algorithm OAGSAA is proposed, and applied to BP neural network, which can effectively analyze and evaluate the security status of the virtual layer. Simulation experiment results show that the method has higher prediction accuracy and convergence speed, and can avoid falling into the local minimum.

1 Introduction

In the process of cloud computing application development, the frequency of security problems also causes enterprises to pay attention to cloud platform security. In the face of the changing security state of cloud platform, it is very important to master the overall security situation of cloud environment in real time by active defense. Cloud platform security situation assessment is an effective and early warning of security issues active defense means.

ENDSLEY[1] put forward the definition of situational awareness for the first time, and think that situational awareness is the perception and understanding of each element or object in the environment and the prediction of the future state under the specific time and space conditions.

In 1999, BASS[2] and others combined the concept of situational awareness with the concept of network security, and gave the definition of network security situational awareness for the first time, that is, to obtain the current network security situation, and to effectively predict the development trend of the future short-term security situation. In the whole process of situational awareness, situational assessment is the key and key link. It collects and analyzes all kinds of evaluation elements, and evaluates the degree of system security threat by establishing mathematical model.

* Corresponding author: 4657547@qq.com

After many years of development, there have been many mature methods and models, but the research on network security situation assessment for cloud platform has not yet formed a more mature model or method.The daily operation of cloud platform will produce massive heterogeneous data. If the security situation assessment model and method for the network are directly applied to the cloud environment, the malicious behavior in the network traffic can only be analyzed. And ignore the host and other parts, especially the virtual machine layer part of the situation.Cloud platform is hierarchical structure, which can be divided into network layer, virtual layer and host layer.At present, most of the methods of cloud platform situation assessment ignore the virtual layer situation, only for the network layer and host layer to extract the situation elements, so the overall situation of the cloud platform is not comprehensive. For example, attackers can masquerade as ordinary users to log on to the cloud platform and control the legitimate virtual machines in the cloud platform to issue DDoS attacks to other virtual machines in the cloud platform, which is a malicious behavior that can not be detected by network traffic. In view of the above problems, this paper mainly studies the situation assessment method for the virtual layer of cloud platform, in order to effectively improve the comprehensiveness of situation perception of cloud platform[3].

2 Related work

Yan Yusheng[4] on the basis of calculation and information security level protection requirements, combined with expert opinion method to establish cloud computing ***IaaS*** security evaluation index system, weight analysis method and analytic hierarchy process combined to process the data, Evaluate according to the evaluation index system.The authority and reliability of the results obtained by the expert opinion method are guaranteed, but the timeliness and expansibility of the method are weak.Huang Ning[5] defines the roughness of cloud network risk system. According to the definition of quantitative dependence and the degree of dependence between risk relationships, the complex cloud network is divided into different simple single systems. After reducing the scale of evaluation, the ***D-S*** evidence theory is used to calculate the situation of each subsystem. Although this method takes into account the multi-layer data in the cloud network, it can not evaluate the behavior from the cloud platform and the security situation of the host.Liu Yuling[6] divided the cloud platform into three levels: virtual machine, physical host and user behavior according to AHP, and calculated the security situation of cloud network environment respectively. The **CSA** obtained by fusion association algorithm fusion calculation is the security situation of the cloud network environment. This method focuses on service security and policy security, and can not evaluate attack behavior in real time. Zhang Chao[7] the hierarchical cloud platform situational awareness model based on cloud platform structure has been improved in synthesis, but the research focus of this model is the optimization of situation prediction method.

3 Improved adaptive genetic simulated annealing algorithm (OAGSAA)

3.1 Introduction to basic algorithms

Genetic Algorithm (GA) is a global search method based on natural selection and genetic principles[8]. It is often used to obtain the optimal solution of the objective function of the optimization problem.GA has strong search ability and high efficiency, so it is widely used to solve various optimization problems.But the biggest disadvantage of genetic algorithm is

precocious, which is often called falling into local optimal value[9].simulated annealing algorithm (Simulated Annealing,SA) is a local search optimization algorithm based on Monte Carlo iterative solution[10]. However, the convergence speed of the algorithm is slow and the oscillation effect may be[11].To solve the above shortcomings, the researchers combined GA with SA to obtain genetic simulated annealing algorithm (Genetic Simulated Annealing Algorithm,GSAA), which not only has GA strong global search ability, but also has good local search ability. It avoids the disadvantage that GA easily fall into local optimization, and improves the overall optimization efficiency of the algorithm.The whole framework of genetic simulated annealing algorithm is similar to that of genetic algorithm. First, the population is initialized randomly, then a new annealing operation is added after the mutation operation, and the new generation population is obtained and the end condition is judged.Since GSAA only fuse the original version of the two algorithms mechanically, and the original GA algorithm has been proved to be unable to completely overcome the disadvantage of falling into local optimum, it needs to be improved[12].Therefore, some detail defects in the algorithm lead to the advantage of GA and SA fusion is not maximized.

BP neural network is a widely used feedforward neural network. The learning process of the classical BP neural network algorithm is based on the gradient descent method, which leads to its shortcomings such as slow convergence speed, low learning accuracy and easy to fall into local minimum value. Especially, the disadvantage of falling into local minimum value is closely related to the result of initial weight threshold selection of BP neural network[13-15].

3.2 An improved OAGSAA algorithm.

3.2.1 Select the optimal individual

The traditional genetic algorithm uses roulette to select the next generation of chromosomes. The core idea of roulette selection is that the probability of individual being selected in the population is proportional to its fitness, and even the chromosome with the highest fitness may not be selected, which is prone to "precocious" phenomenon and fall into local optimum. In order to solve the above problems, the individuals with the highest fitness in each generation are directly inherited to the next generation, and the remaining individuals are selected to enter the next generation according to a certain probability. P probability formula is as follows:

$$P=\frac{fit_i}{\sum_{i=1}^{G} fit_i - fit_{max}} \tag{1}$$

This can ensure that the optimal individual must inherit to the next generation, and reduce the probability of precocious phenomenon.

3.2.2 Adaptive probability of crossover variation

The crossover probability ***Pc*** and mutation probability ***Pm*** are generally set as fixed values in traditional genetic algorithms, but too large or too small setting will affect the effect of the algorithm. If the set probability is not appropriate, it will increase the possibility of the algorithm falling into local extremum or reduce the efficiency of the algorithm. The paper adopts adaptive method to determine Pc and Pm can overcome this problem.The formula of

adaptive crossover probability Pc and mutation probability P and mutation probability Pm as follows:

$$P_c = \begin{cases} k_1 \sin\left(\frac{\pi fit_{avg}}{2fit_{max}}\right) & \frac{fit_{avg}}{fit_{max}} \geq \frac{1}{2} \\ k_1\left(1 - \sin\left(\frac{\pi fit_{avg}}{2fit_{max}}\right)\right) & \frac{fit_{avg}}{fit_{max}} < \frac{1}{2} \end{cases} \quad (2)$$

$$P_m = \begin{cases} k_2\left(1 - \sin\left(\frac{\pi fit_{avg}}{2fit_{max}}\right)\right) & \frac{fit_{avg}}{fit_{max}} < \frac{1}{2} \\ k_2 \sin\left(\frac{\pi fit_{avg}}{2fit_{max}}\right) & \frac{fit_{avg}}{fit_{max}} \geq \frac{1}{2} \end{cases} \quad (3)$$

After each crossover operation and mutation operation, the feasibility of the two chromosomes after crossover must be detected. If the two chromosomes are abnormal, the operation result is not feasible and the crossover or mutation operation needs to be carried out again.

3.2.3 Metropolis guidelines for increasing acceptance rates

A Metropolis criterion is used to determine whether a new state is accepted in a simulated annealing algorithm.When the probability ***P*** is larger than the random number produced in the interval [0,1], the accepted state is j a new state, whereas the original state is retained.The traditional Metropolis criterion is adopted in the GSAA, which is accepted only when the new state is good enough, which will abandon too many individuals with low fitness and reduce the diversity of the population.:Suppose that when the temperature is T, the current state changes from i to a new state j, and the corresponding internal energy is E_i and E_j,respectively When $E_i<E_j$, the accepted j is the current state,If $E_i \geq E_j$, calculate the probability:P =exp[- (Ej -E i)/KT], ***K*** is a Boltzmann constant.When the probability P is larger than the random number produced in the interval [0,1], the accepted state is E_j a new state, whereas the original state is retained,on the contrary, the original state is still retained.The traditional Metropolis criterion is adopted in the GSAA, which is accepted only when the new state is good enough, which will abandon too many individuals with low fitness and reduce the diversity of the population.To improve the acceptance rate of individuals with low fitness so as not to fall into premature maturity, this paper proposes OAGSAA algorithm to improve the Metropolis criteria: On the assumption that the fitness of the new individual after the selection, crossover and mutation operation is f g(′), the original individual g fitness is f (g), the probability of accepting the new individual to replace the old individual is:

$$P=\begin{cases} 1 & f(g') \geq f(g\ \ ) \\ exp\left(-\frac{f(g')-f(g\ \ )}{T_c}\right) & f(g') < f(g\ \ ) \end{cases} \quad (4)$$

where T_c =T_0/N_c , T_0 is the initial annealing temperature and N_c is the current evolutionary algebra N_c <N.Optimized metropolis improves the acceptance rate for individuals with lower fitness and preserves the diversity of populations and the effectiveness of annealing processes as much as possible.

3.3 A OAGSAA-BP virtual layer situation assessment algorithm for cloud platform.

BP algorithm has good data classification ability and has been applied in network situational awareness system, but its slow convergence and easy to fall into local minimum have a certain impact on the effect of situation assessment.Therefore, the optimized OAGSAA algorithm is used to optimize BP neural network, that is, OAGSAA-BP algorithm.Therefore, the optimized OAGSAA algorithm is used to optimize BP neural network, that is, OAGSAA-BP algorithm. The algorithm flow is shown in figure 1.

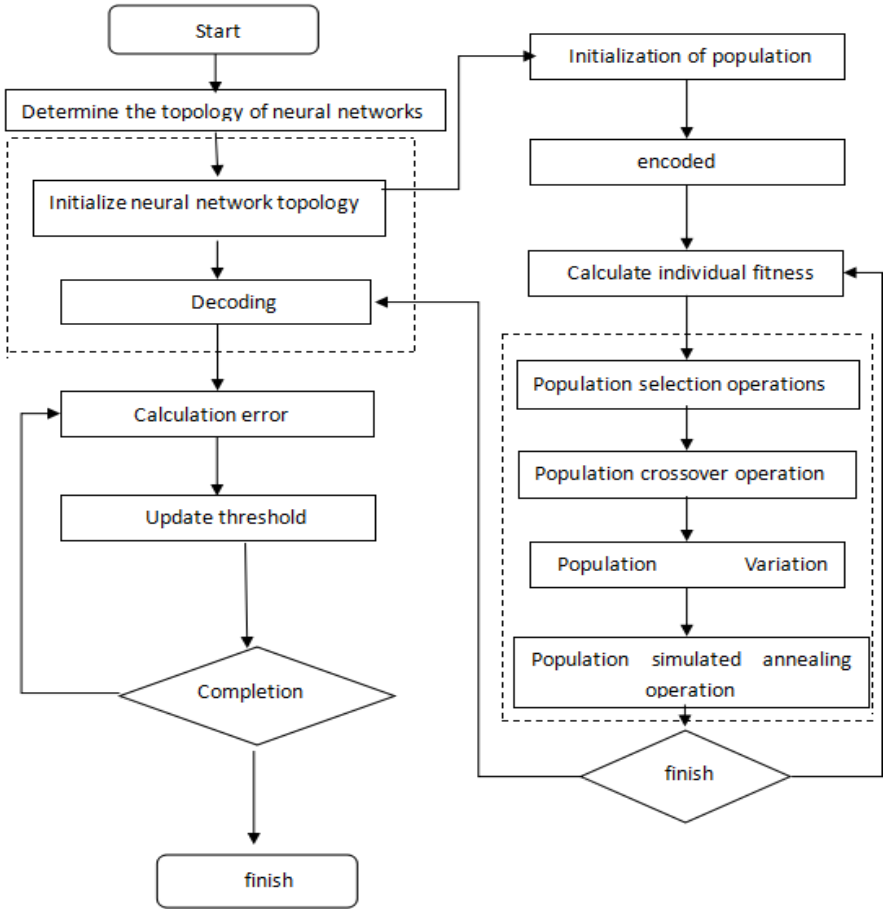

Fig. 1. OAGSAA-BP Algorithm flow.

The algorithm flow is described as follows :

1) the topology of the neural network is determined according to the collected samples and the algorithm is initialized.The floating-point coding is used to encode the parameters and initialize the population size and other related parameters, and the current evolutionary algebra is set to 1.Chromosome length L calculation formula is:

$$L = I_i H_j + H_j + H_j O_k + O_k \tag{5}$$

The I_i is the number of input layer nodes and the H_j O_k is the number of output layer nodes.

2) Calculate the individual fitness according to the fitness function and compare it. The fitness value is proportional to the individual's merits and demerits, and the optimal individual with the highest fitness is recorded.The fitness function is:

$$Fit = \frac{1}{\sum_{i=1}^{N}(y_i' - y_i)^2} \tag{6}$$

3) The population selection, crossover, mutation operation to produce a new population.

4) The new population is annealed according to the optimized Metropolis criterion. If the new population is accepted, the original population is replaced by the new population and the relevant parameters are updated.

5) Evolutionary algebra plus 1, $N_c = N_c + 1$.

6) If $Nc < N$ and $T_c \geq T_{min}, T_c = T_0/N_c$ and go to cycle step 2); if $N_c \geq N$ or $T_c \geq T_{min}$, the genetic process or annealing process ends, returns the optimal individual and goes to step 7.

7) Decoding the optimal individual to obtain the optimal initial weight threshold of the neural network, starting to BP the neural network cycle process, until the actual output error meets the accuracy requirements, the algorithm ends.

4 Experimental simulation and analysis.

This paper first defines the security situation elements needed for cloud platform virtual layer situational awareness, as shown in Table 1.

Table 1. Security situational awareness requirements for virtual layer VDSI.

Serial number	Project	Meaning
1	time	Time
2	mac-s	Source mac address
3	mac-d	Aim mac Address
4	protocol	Agreement
5	IP-s	Address IP source virtual machine
6	IP-d	Destination Virtual Machine IP Address
7	port-s	Source Virtual Machine Port
8	port-d	Purpose Virtual Machine Port
9	number	Number of data
10	length	Packet length
11	TTL	Survival time
12	exe_num	Total number of virtual machine processes
13	exe_sleep	Number of virtual machine dormant processes
14	exe_run	Number of virtual machine running processes
15	exe_zom	Number of virtual machine dead processes
16	exe_stop	Number of virtual machine termination processes
17	hidden_proess	Number of virtual machine hidden processes
18	thread_num	Number of virtual machine threads
19	Unlinked_DLLs	DLL not linked
20	Loaded_DLLs	DLL the process has loaded
21	dll_num	DLL files changed
22	ART	Average response time
23	Loss Tolerance	Loss rate
24	Protection_software	Does the virtual machine turn on protection

In order to better simulate the real cloud computing environment, this paper uses the honeynet data set published by the honeynet project team as the experimental test data[16]. There are many systems, servers, virtual machines and so on in Honeynet, and its structure

is very similar to that of cloud platform, which can simulate the running data of cloud platform.For converting a large number of different dimensional data into dimensionless data, the corresponding Onehot coding and normalization of data sets are also needed.The normalized formula is as follows:

$$x_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} (i = 1,2,\dots,n) \tag{7}$$

Because of the large amount of data in the original data set, in order to make the experiment operation convenient and the experimental results obvious, the data set is extracted according to the original proportion to form different subsets of the test data[17].Data set $D_1D_2D_3$, for all states containing raw data, data sets of 1 normal state and 6 attacked states, Each dataset contains 500 pieces of data. Data set $D_4D_5D_6$, a data set containing only one of Normal, Ping, DNS states, Each dataset contains 100 pieces of data. Each data set is tested with a random selection of 90% of the training data, The remaining 10% is the test data. Using BP algorithm, GA-BP algorithm and OAGSAA-BP algorithm respectively Line test, the accuracy of the classification results is shown in figure 2

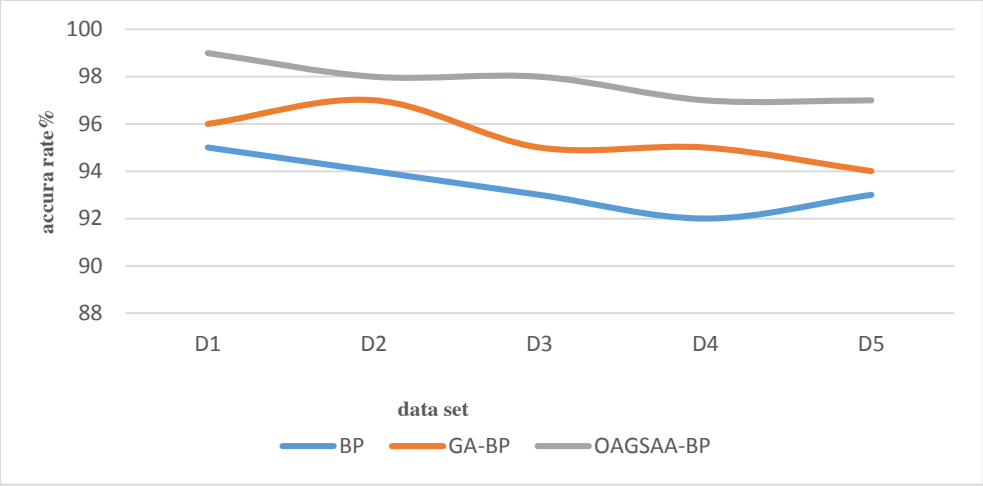

Fig. 2. Algorithm accuracy.

It is not difficult to see from figure 2 that the accuracy of the BP algorithm fluctuates the most, followed by the GA-BP algorithm, and the OAGSAA-BP algorithm performs relatively well, both the multi-state data set and the single-state data set can have a more stable accuracy. The average accuracy of the three algorithms is 95.49%,96.41% and 97.61% respectively.

The slow convergence rate is an obvious disadvantage in the practical application of BP algorithm. One advantage of OAGSAA-BP algorithm is that it can obviously improve its iteration rate. Use BP algorithm, GA-BP algorithm and OAGSAA-BP algorithm to record the running time respectively, and the running time comparison is shown in figure 3.

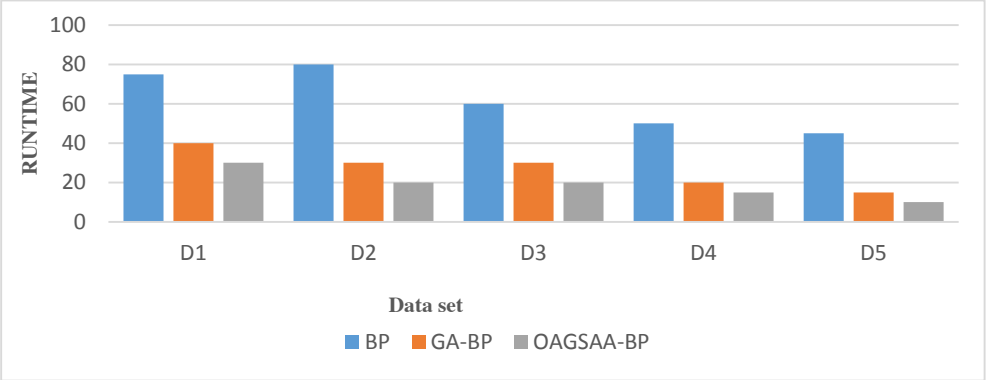

Fig. 3. Running time of algorithm.

It is not difficult to see that the optimized algorithm does solve the problem of slow convergence, and the running time of each data set is the shortest OAGSAA-BP algorithm. The comparison between the D1-D3 data set and the D4-D5 data set shows that when the amount of data increases exponentially, the convergence rate of the BP algorithm decreases, followed by the GA-BP algorithm, while the convergence rate of the OAGSAA-BP algorithm changes the least. It can be seen that the efficiency of the algorithm is relatively stable.

Mean square error (MSE) and mean absolute error (MAE) are commonly used to evaluate the overall error of the algorithm. The error pairs D1 the data set are shown in Table 2.

Table 2. Comparison *D1* error between data sets.

Algorithm	Evolutionary generation	MSE	MAE	accuracy rate
GA-BP	36	0.396	0.422	96.6%
OAGSAA-BP	11	0.213	0.287	97.0%

With figure 3 and Table 2, it can be concluded that the OAGSAA-BP algorithm can effectively improve the convergence speed of the original algorithm, and with the change of data volume, the running time will not increase exponentially, and the MSE and MAE values are smaller than the GA-BP algorithm.

The purpose of situation assessment on the virtual layer of cloud platform is to integrate and evaluate the situation elements collected and get the current situation to prepare for the situation prediction stage. Figure 4 is a comparison of the average accuracy of the data set $D_1D_2D_3$,7 states.

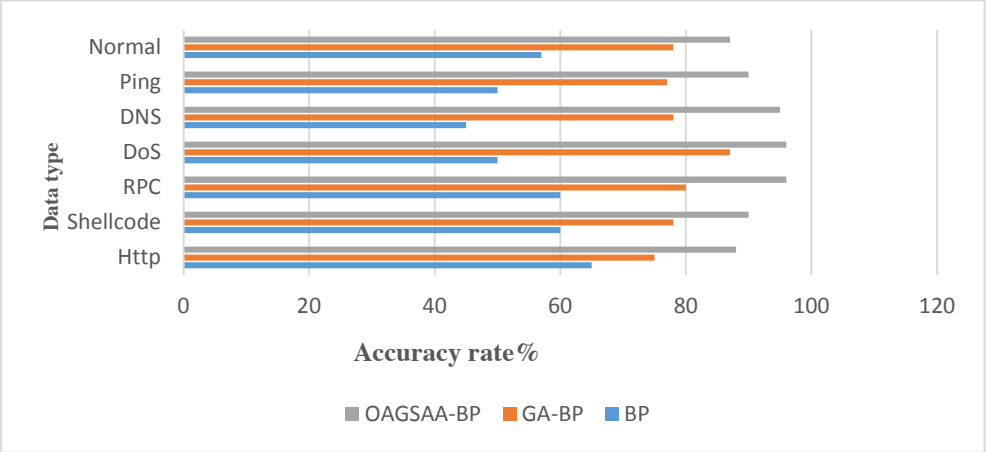

Fig. 4. Comparison of accuracy.

Compared with the two algorithms before the improvement, the OAGSAA-BP algorithm has higher accuracy for the evaluation results of various situation types, and does not fluctuate because of the small number of certain kinds of data. Tables 3 and 4 divide the security level and the cloud platform virtual layer security level.

Table 3. Classification of state security levels.

	Level of threat	Corresponding situation values
Normal	0	0
Ping	1	0.2
DNS	2	0.5
DoS	2	05
RPC	2	0.5
Shellcode	3	0.8
Http	3	0.8

Table 4. Security Level of Virtual Layer of Cloud Platform.

Virtual Situation Value of Cloud Platform	Corresponding security level
0-0.2	security
0.2-0.5	Mild dangerous
0.5-0.8	Moderate dangerous
0.8-1	High dangerous

When the virtual layer of the cloud platform is in a Normal state, the current situation value of the virtual layer of the cloud platform is 0; if the virtual layer of the cloud platform is only attacked by the DNS, the current situation value of the virtual layer of the cloud platform is 0.5; if the virtual layer of the cloud platform is subjected to a variety of Attack, according to formula (8) weighted calculation, the current cloud platform virtual layer situation value S obtained.

$$S = \frac{n\sum_{i=1}^{n} S_i w_i}{\sum_{i=1}^{n} S_i} \tag{8}$$

S_i is the situation value corresponding to each attack, wi the threat degree corresponding to the attack. S $\in$[0,1], when the S is greater than 1, the current state is highly dangerous and the S value is 1.

Eight time points were randomly selected for situation value correspondence test. The attack and situation values at each time point are shown in Table 5 and figure 5.

Table 5. Attacks at test points.

Time point	Number of attack	Type of attack
T1	1	DoS
T2	0	None
T3	1	Ping
T4	0	None
T5	2	Ping,Shellcode
T6	1	Http
T7	3	Ping,DNS,Http
T8	2	Ping,DoS

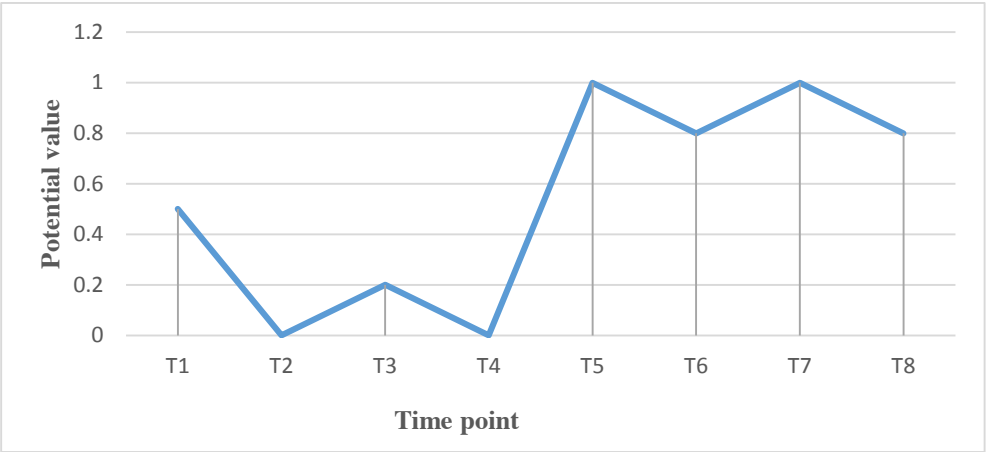

Fig. 5. Potential values at test time points.

The t_2 and t_4 in figure 5 are safe, the t_3 is mild, the t_1 is moderate.As can be seen, after the security elements of the virtual layer of the cloud platform are accurately classified by the OAGSAA-BP algorithm, the corresponding situation values can accurately reflect the security state of the current virtual layer of the cloud platform, and lay the foundation for the next step of cloud platform situation prediction.

5 Conclusion

This paper proposes an improved adaptive genetic simulated annealing algorithm for virtual layer based on the weak situation perception of virtual layer by the existing cloud computing situational awareness model OAGSAA,and combines with BP neural network to evaluate the dynamic security state of virtual layer.The algorithm improves the global search ability of genetic algorithm by defining adaptive crossover, mutation probability parameters, optimization Metropolis criteria, and is used to optimize the BP neural network algorithm to improve the convergence rate and global optimization effect of the algorithm.From the simulation results, the convergence speed of the OAGSAA-BP algorithm is faster, the classification accuracy is higher, the overall performance of the algorithm is more stable, and the dynamic security index of the virtual layer of the cloud platform can be evaluated more accurately and effectively, which provides an effective support for the overall security situation assessment of the cloud platform.

References

1. ENDSLEY M R. Toward a Theory of Situation Awareness in Dynamic Systems[J]. Human Factors, 1995, 37(1): 32-64.
2. BASS T, GRUBER D. AGlimpse into the Future of ID[J]. The Magazine of USENIX&SAGE, 1999, 24(3): 40-49.
3. CHEN Xiuzhen, ZHENG Qinghua, GUAN Xiaohong, et al.Quantitative Hierarchical Threat Evaluation Model for Network Security[J]. Journal of Software, 2006(4): 885-897.
4. YAN Yusheng. Security Assessment Research for Cloud IaaS Service Based on Classified Protection[D]. Beijing: Beijing Jiaotong University, 2016.
5. HUANG Ning. Research on Network Security Situational Awareness Technology in Cloud Computing Environment[D]. Xi'an: Xi'an Polytechnic University, 2018.
6. LIU Yuling, JIA Xuefei, YAN Yan, et al. Cloud Security Situation Awareness Method Based on Gray Theory[C]//The 15th Research Institute of China Electronics Technology Group Corporation. National Information Security Level Protection Technology Conference, July-3, 2015, Beijing, China. Beijing: China Electronics Technology Group Corporation, 2015: 21-24.
7. ZHANG Chao. Research and Analysis of Cloud Computing Network Security Situation Assessment[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
8. KEL-SHORBAGYMA,AYOUBAY, EL-DESOKYIM, et,al. A Novel Genetic Algorithm Based K-means Algorithm for Cluster Analysis[M]. Heidelberg: Springer-Verlag, 2018.
9. LI K, JIA L, SHI X. An Efficient Hybridized Genetic Algorithm[C]//IEEE. International Conference of Safety Produce Informatization (IICSPI),November 23-25, 2018, Chongqing, China. New York: IEEE, 2018:118-121.
10. METROPOUSN. Equations of State Calculations by Fast Computing Machines[J]. The Journal of Chemical Physics, 1953(21): 1087-1091.
11. CHEN Jun, HE Li, QUAN Yi, et al. Application of BP Neural Networks Based on Genetic Simulated Annealing Algorithm for Shortterm Electricity Price Forecasting[C]//IEEE.International Conference on Advances in Electrical Engineering (ICAEE), January 9-11, 2014, Vellore,India. New York: IEEE, 2014: 1-6.
12. YUAN Yongliang, WANG Guohu. Self-adaptive Genetic Algorithm for Bucket Wheel Reclaimer Real-Parameter Optimization[J]. IEEE Access, 2019: 47762-47768.
13. Ding Qunyan. Improved BP Neural Network Controller Based on GA Optimization[C]//IEEE. International Conference on Smart Grid and Electrical Automation(ICSGEA), May 27-28, 2017, Changsha, China.New York: IEEE Computer Society, 2017: 251-254.
14. Yuan Zhenbing, Guo Shuli, Han Lina. Disease Diagnostic Prediction Model Based on Improved Hybrid CAPSO-BP Algorithm[C]//Technical Committee on Control Theory (TCCT). 2017 36th Chinese Control Conference, July 26-28, 2017, Dalian, China. CCC, 2017: 3960-3965.
15. KANG Zhou, QU Zhiyi. Application of BP Neural Network Optimized by Genetic Simulated Annealing Algorithm to Prediction of Air Quality Index in Lanzhou[C]//IEEE.International onference on Computational Intelligence and Applications, September 8-11, 2017, Beijing, China.IEEE, 2017: 155-160.