Research and simulation of network security situation prediction algorithm

Zhengmao Li¹, Tingting Ma¹, Yanling Zhou¹ and Xianhui Wang¹

¹School of Artificial Intelligence and Big Data, Hefei University, Hefei, Anhui Province, China, 230601

Corresponding author and e-mail: Zhengmao Li, lizm@hfuu.edu.cn

Abstract. Monitoring and predicting the security state of network environment is the key to solve network security. On the basis of studying the principle of network security situation awareness, a network security situation prediction model based on artificial immunity and Kalman algorithm is designed according to the artificial immunity evaluation model. First, Wireshark software was used to analyse DARPA99 dataset. Then, the situation value was obtained by using the artificial immune evaluation model. In the Curver Fitting Tool, the Kalman algorithm is written to achieve the optimal prediction of the situation value. The test results show that the model has a good prediction effect.

1. Introduction

The prosperity of the Internet has provided a powerful driving force for social and economic development, but it has also brought great challenges to cyber security, posing a serious threat to social and national security. Network security situation prediction is a hot research topic at present. It’s found out there are risks of network environment in time, we need to prevent and deal with them in advance, so as to ensure the security of network environment.

Endsley proposed the concept of network security situational awareness in 1988, and proposed the network security situational awareness model in 1999 [1]. Yu Jian et al. proposed a safety situation awareness prediction method based on grey correlation [2]. Fan Jiulun and Cheng Jiagen respectively proposed network security situation prediction method based on RBF neural network [3][4]. However, the RBF algorithm needs to use the samples of the latest period to train the parameters in the first place, so it cannot make timely and effective prediction for the sudden attack behavior of the network. Qin Lina put forward a network Security Situation Prediction Based on Artificial Immune Algorithm. Using the similarity between network security situation and artificial immunity, antibody concentration was used to optimize network parameters and network output results [5]. Xiang Xixi et al. applied Kalman algorithm to network security situation assessment and prediction model [6].

In this paper, by studying the network security situation awareness model, the artificial immunology is applied to the network security situation assessment, and the Kalman algorithm is used to construct the linear equation of state and observation equation. By calculating the covariance, the gain coefficient is obtained, so as to obtain the optimal predicted value. Then, according to the network security situation at time “t”, the network security situation at time “t+1” is calculated, and the network security situation is predicted.
2. Network security situation prediction architecture design

Network security situational prediction refers to the evaluation of current network security state and prediction of future development trend based on network data obtained from known network environment, which is used to make security decisions and improve network defense capability.

2.1. Construction of network security situation model

It mainly includes three modules: extraction of network security situation factors, situation assessment and situation prediction.

Extraction of network security situation factors is the premise of security situation assessment. The factors that have influence on network security such as attack intensity, network traffic and traffic change rate are extracted.

Network security situation assessment refers to the analysis and processing of the obtained network security situation factors. A numerical value representing the threat degree of the current network system under attack is obtained by calculating the appropriate mathematical model.

Network security situation prediction refers to finding the relationship between current situation value, future situation value and other influencing factors according to the situation value information obtained from situation assessment, obtaining some calculation equation, and then combining with the influencing factors to obtain a predicted situation value.

2.2. Model Functional module design

Specifically in this paper, the function module of the model mainly includes three parts: obtaining the attack times, calculating the situation value and the situation value prediction.

![Figure 1. Function block diagram.](image)

2.2.1. Obtaining the attack times. In this paper, attack intensity is chosen as the network security situation factor, and other factors are dealt with in similar ways. First, Wireshark's filter engine is used to filter out packets that users need. And its own I/O graphic tools are applied to displaying the flow of the capture package file. For example, when detecting SYN-flood attack, SYN message is at a relatively low level under normal traffic condition. When a large number of SYN packets are detected, it can be judged that the system is under SYN-flood attack.

2.2.2. Calculation of the situation value. The function of calculating situation value belongs to the security situation assessment module, and the artificial immunity assessment model was used in this paper.

The relationship between host and network attack is analogous to that between antibody of human immune system and pathogen invasion. When pathogen invasion occurs, the number of antibodies in the body increases, and when the attack decreases, the number of antibodies gradually decreases. Therefore, the network security situation assessment method based on artificial immunity can well
reflect the change of host security degree when attacked. The calculation of the situation value in this paper refers to the calculation of the situation value when the whole system contains multiple servers under certain attack. Under type J attack, the system posture value is calculated as shown in Formula (1) [7].

\[ R_j^i = 1 - \frac{1}{1 + \ln(\alpha_j \sum \beta_i n_{ij} - x_i | + 1)} \]  

(1)

\( N_{ij} \) is the number of antibodies to detect type j attacks on host I, \( \alpha_j \) is the danger of type J attack, and \( \beta_i \) is the importance of host i. According to the formula, the greater the situation value is, the greater the risk of the current system is, and the more dangerous the system is.

2.2.3. Situation value prediction. Kalman algorithm is used to make the prediction. Kalman filter is an algorithm that utilizes linear system state equation to estimate the system state optimally through the system input and output observation data. Attack intensity has the greatest correlation with situation value. For the convenience of the experiment, only the control variables affecting attack intensity on situation value are selected. In the artificial immune model, the risk value was calculated by antibody concentration, based on which the Kalman filter prediction model could be established. Combining the network security situation value of the previous period with the change value of the current attack intensity, a binary one-order function model is established with the network security situation value of the current period. The regression model is as follows:

\[ X_{k+1} = AX_k + BU_k + W_k \]  

(2)

\[ Y_{k+1} = HZ_{k+1} + V_k \]  

(3)

\( X_{k+1} \) represents the network situation value at time k+1, \( X_k \) represents the network situation value at time k, \( U_k \) represents the change of attack intensity at time k, \( W_k \) represents the state transition error at time k, \( Z_{k+1} \) represents the calculated risk value at time k+1, and \( Y_{k+1} \) represents the observed value of situation value at time k+1. \( A \) is the state vector, \( B \) is the control vector, \( H \) is the observation transfer vector, \( W_k \) is the process noise, and \( V_k \) is the observation noise.

The prediction principle based on Kalman algorithm is to find a linear description equation that represents the relationship between the situation value at the moment k and the situation value at the previous moment and the control variable, and to list the observation equation. It can be seen from the real environment that both equations inevitably contain certain errors. The Kalman gain coefficient k with the minimum covariance can be obtained through Kalman filtering algorithm. The predicted value is equal to the desired value of the state equation plus k times the difference between the state value and the observed value.

Due to two equations inevitably contains a certain error, the Kalman gain coefficient k with the minimum covariance can be obtained through Kalman filtering algorithm. Predicted value is equal to the state equation desired value plus k times the difference between the state value and the observed value, and at this time are to satisfy the desires of the optimal estimate.

3. Simulation and result analysis
In Windows 7.0, Wireshark, Matlab, Curver Fitting Tool and other tools were used to implement the algorithm.

3.1. Module of obtaining network security situation factors
Wireshark software is used to open part of Darpa99 data set, and I/O chart tool is used to analyze the data set. Display Filter is set as TCP. Flags ==0x02, Y Axis is set as Packets. The filtered attack packets are of type SYN. The analysis results are shown in Figure 2. Data packets of 30 consecutive periods including the peak value are selected to record and stored in the file shuju.mat.
3.2. Module of network security situation assessment

Equation (1) was used for situation assessment to calculate the situation value, and the risk of SYN-FLOOD attack was set as 0.8. In the environment, it was assumed that two servers were attacked, and the weights were set as 0.3 and 0.5 respectively. The calculated situation value is normalized to obtain the image display, as shown in Figure 3.

According to the number of captured attack packets, it can be seen that the number of 0-24s network attack packets is at a normal level, while the number of 25s attack packets is significantly increased, and it returns to the normal level after 26s. Combined with the situation diagram, the 0-24s situation value is lower than 0.5, and the situation value increases to 0.95 around 25s, indicating that the system is more dangerous, and tends to around 0.3 at 26s. It can be seen that the situation value evaluation method is more reasonable and can reflect that the situation value increases with the increase of attack intensity.
3.3. Module of network security situation prediction
The equation of state was created according to the Kalman prediction model, and the parameters A and B of the equation of state were calculated by using the Curver Fitting Tool. The fitting equation is solved as shown in Figure 4.

![Fitting equation diagram](https://via.placeholder.com/150)

Figure 4. Fitting equation diagram.

The situation value T3 of 1-29s and the normalized value U2 of 1-29s change in attack times were taken as the independent variable, and the situation value T4 of 2-30s was taken as the dependent variable to construct a multivariate function equation. As can be seen from the figure, the parameters in the equation of state obtained were A=0.8147 and B=0.832. The regression prediction equation can be obtained as shown in Formula (4).

$$X_{k+1} = 0.8147X_k + 0.832U_k$$

(4)

A predicted value can be obtained by combining the obtained equation with the normalized value of the current attack intensity change. The predicted value is compared with the situation value obtained by the situation assessment, and the result is shown in Figure 5. The green line represents the predicted result based on the equation of state, while the red line represents a state value obtained by observation and calculation. It can be seen that the equation described above can basically predict the general trend of state value.

Since both the observation equation prediction and the state equation prediction contain errors and are only rough predictions of the state value, the Kalman prediction algorithm can realize the optimal estimation based on the two equations.
The prediction results are shown in Figure 6. The blue line represents the predicted value obtained after the application of Kalman algorithm. Kalman algorithm is a Kalman gain obtained by integrating two equations to obtain the minimum full football covariance, so as to obtain the optimal predicted value. It can be seen from the result figure that the blue line is between the red line and the green line, and closer to the red line, indicating that the observed values are more likely to reflect the change of the system's situation value, and also indicating that the Kalman prediction model has better prediction effect compared with the ordinary linear regression prediction model.

4. Conclusions
This paper mainly studies the principle of network security situation awareness, and designs a network security situation prediction model compared with the evaluation model of artificial immunity based on artificial immunity and Kalman algorithm. Finally, the simulation is carried out. The results show that Kalman algorithm achieves the expected prediction effect. However, only one SYN-Flood attack analysis was selected for obtaining the security situation value in this experiment, while it were too limited for the factors considered, and was insufficient for the dynamic nature of prediction. These all will be the focus of the next step.

Acknowledgement
This work was financially supported by the Quality Engineering of Higher Education of An Hui Province (No. 2018jyxm1105), The Quality Engineering of Higher Education of Hefei University (No.2018hfppkc06, No.2018hfsjzd08, No.2018hfxgk01, and No. 2019xbkc005).

References
[1] Endsley M R. Situation awareness global assessment technique (SAGAT)//Aerospace and Electronics Conference, 1988. NAECON 1988., Proceedings of the IEEE 1988 National. IEEE, 1988: 789-795.
[2] Yu Jian, Lin Zhixing, Xie Bin. Research and Exploration in Laboratory, 2019, 038(002):31-35,70. (In Chinese)
[3] Fan Jiulun, Wu Peng. Network Security Situation Prediction Method Based on RBF Neural Network [J]. Journal of Xi ‘an University of Posts and Telecommunications, 2017, 022(002):7-11. (In Chinese)
[4] Cheng Jiagen, Qi Zhenghua, Chen Tianfeng. Network Security Situation Awareness Based on RBF Neural Network [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition),2019,39(04):88-95. (In Chinese)
[5] Qin Lina. Network Security Situation Prediction Based on Artificial Immune Algorithm [J].
Journal of Changchun Institute of Technology (Natural Science Edition). 2018, 19(01). (In Chinese)

[6] Xiang Xixi, Huang Hongguang, Li Yudong. Network Security Situation Prediction Based on Kalman Algorithm [J]. Computer Simulation, 2010, 027(012):113-116. (In Chinese)

[7] Ke Yi Zhang; Cheng Long Hao; Jie Yang; Jia Qi Fan; Bo Yu Wang. Research of Network Security Situation Prediction Based on Kalman Algorithm [J]. Advanced Materials Research Vol. 933(2014) pp 714-719.