Optimization Design of Large-Scale Network Security Situation Composite Prediction System

Jinbao Shan(✉) and Shenggang Wu

College of Information Technology and Art Design,
Shandong Institute of Commerce and Technology, Jinan, China
sammyel980@yeah.net

Abstract. Because the traditional network security situation compound prediction system cannot overcome the defects of SVM algorithm, the accuracy of extraction results is low. For this reason, a large-scale network security situation compound prediction system is designed. Through data normalization process to optimize the SVM algorithm, to optimize the forecasting calculation module, to provide data base system frame structure, system frame structure can be divided into security situational composite sensing module, situational composite evaluation module and situational composite prediction module, synergy is derived using multiple module network security situational values when attacked, to implement network security situation prediction to complete the system design. Simulation application environment design compared the experimental results show that compared with the traditional prediction system of the proposed system under the same data to forecast, the accuracy of predicted results by 65%, and the operation is very stable.

Keywords: Large-Scale · Network security situation compound forecast · Forecast result · Accuracy

1 Introduction

With the deepening of information technology, the Internet is becoming the critical, information infrastructure of the country, cyber security concerns the fundamental interests of the country and society. In recent years, the global internet is frequently attacked, lead to growing network security issues, the security of important information systems is seriously threatened. In order to deal with the challenges of network security, security protection and management systems such as VPN, IDS, antivirus systems, identity authentication, data encryption, and security audit have been widely used, however, the accuracy of the prediction results of the traditional network security situation composite forecasting system is lower, this paper uses the optimized SVM algorithm as the basis for establishing the prediction system, divide the system into three parts: perception, assessment and prediction, calculate the security posture of the network when it is attacked by the system. The experimental results show that the system designed in this paper is more suitable for the complex prediction of large-scale network security situation than the traditional system.

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 419–425, 2019.
https://doi.org/10.1007/978-3-030-36405-2_42

2 Optimization Design of Large-Scale Network Security Situation Composite Prediction System

2.1 Predictive Calculation Module Design

Because the range of network security situation changes is relatively large, it has an adverse effect on the training speed of SVM. The reconstructed data is input into the model for learning and normalized, and the specific normalization is [1, 2]:

$$x_i^{'} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Among them, x_i is the original value, $x_i^{'}$ is the normalized value, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values respectively.

After normalizing the data, design the system calculation model.

Let there be a total of n network security situation learning samples $\{ \mathrm{x_i}, \mathrm{y_i} \}$, $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$, x_i is input, y_i is the output expectation where the input is the expected value of the output and the SVM regression equation is [3]

$$f(x) = w \times \varphi(x) + b \tag{2}$$

$$\varphi: R^n \to G, w \in G \tag{3}$$

In the formula, w represents the weight vector, b represents the offset vector

Use the optimization function to optimize the target value, that is:

$$\min J = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{\mathrm{n}} (\xi_i^* + \xi_i) \tag{4}$$

The constraints are as follows:

$$\begin{aligned} &y_i - w \times \varphi(x) - b \le \varepsilon + \xi_i \\ &w \times \varphi(x) + b - y_i \le \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \ge 0, i = 1, 2, \ldots, n \end{aligned} \tag{5}$$

Among them, ξ_i, ξ_i^* is expressed as a relaxation factor, and C is a penalty factor [4].

By introducing Lagrange multiplier, we can get:

$$\begin{aligned} L(w, b, \xi, \xi^*, \alpha, \alpha^*, y, y^*) = &\frac{1}{2}\|w\| + C\sum_{i=1}^{n} (\xi_i + \xi_i^*) - \sum_{i=1}^{n} \alpha_i(\xi_i + \varepsilon - y_i + f(x_i)) \\ &- \sum_{i=1}^{n} \alpha_i^*(\xi_i + \varepsilon - y_i + f(x_i)) - \sum_{i=1}^{n} (\xi_i \gamma_i - \xi_i^* \gamma_i^*) \end{aligned} \tag{6}$$

Among them, α_i and α_i^* are Lagrange multipliers, and ε is a parameter of insensitive loss function.

The SVM regression expression is thus obtained as [5–7]:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \times (\varphi(x_i), \varphi(x)) + b \tag{7}$$

For non-linear regression prediction problems, to prevent dimensional disaster problems, use kernel function $k(x_i, x)$ instead of $(\varphi(x_i), \varphi(x))$, to prevent dimension disasters, that is:

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x_i, x) + b \tag{8}$$

In summary, the calculation model used by the large-scale network security situation composite forecasting system is:

$$\begin{aligned} f(x) &= \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) \times (\varphi(x_i), \varphi(x)) + b \\ &= \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x_i, x) + b \end{aligned} \tag{9}$$

2.2 System Framework Design

The large-scale real-time network security situation composite forecasting system collects the information of each node of the network in real time, and carries out security situation composite sensing, situational compound assessment and situational compound prediction for the entire network. The framework of the system is shown in Fig. 1. It is mainly composed of data collection, security situation composite analysis, security situation compound assessment, security situation composite forecast and assessment database [8, 9].

Data collection includes two aspects. The first is the real-time collection of IDS alarm logs. It contains a large number of network attack information and is an important data source for the complex assessment of network security posture. The second is the network node information, which is calculated due to network risk assessment. The security situation is more theoretical and should be corrected in light of the real-time performance of network nodes. Security situational complex perception mainly includes attack information extraction and threat identification. Because there are many IDS alarms, such as Snort basic alarms, there are more than 8000 kinds. Therefore, it is necessary to extract and classify attack behavior according to the threat degree of the alarm, which can reduce the assessment and the complexity of the forecast. The security situation compound assessment evaluates the network security situation based on the attack information. Security situation compound prediction uses

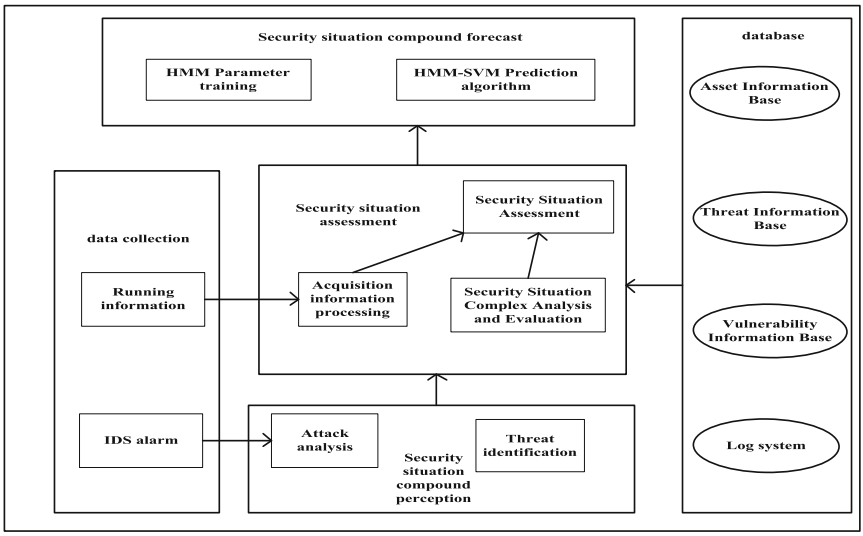

Fig. 1. Large-scale network security situation composite forecasting system framework

hidden Markov model for situational composite prediction. Firstly, the HMM parameters are trained according to the results of past evaluations. Then, the HMM-SVM prediction model is used to predict the next state of the network. The evaluation database contains an asset information base, threat information base, vulnerability information base, and log system. Host information scanning program is used to obtain host configuration information, including application programs, operating systems, vulnerability scanners to scan out port numbers, vulnerability numbers, and possible Consequences, etc. [10].

2.3 Realize the Network Security Situation Compound Prediction

Set the delay time of the network security situation composite data transmission, use the test method to embed dimension into the situation data, and determine the embedding dimension (assumed to be n). At this point, the support vector machine has $n - 1$ input variables and one input variable. Based on the delay time and the embedding dimension, the data of the current reaction network security situation is reconstructed, and then the training samples and test samples of the support vector machine are generated. The data representing the training sample is input into a support vector machine for learning, and is optimized using a calculation model, and the parameter values of the algorithm are set. When the optimal parameters of ε, C, and φ are brought into the prediction model, then the prediction model becomes the optimal network security situation prediction model. Through this model, the security situation value of the network when it is attacked can be calculated. The model is used to predict and analyze the previous generations of security situation values, and the characteristic data of the trend data are depicted. By analyzing the curves, it can be seen whether the

established network security situation prediction system can effectively predict the safety monitoring data and whether the relevant prediction data has higher accuracy.

3 Analysis of the Experiment

3.1 Experimental Data

In order to verify the correctness and effectiveness of the large-scale network security situation composite forecasting system designed in this paper, simulation experiments were done in this paper. A network includes three attacked nodes, namely an HTTP server, a mail server, and an FTP server. It is assumed that each node has a weight of 0.5, 0.3, and 0.2 in the network. This paper simulates Attacker Attacker launching scan attack (i.e. attack 1), buffer overflow attack (i.e. attack 2) and TCP-SYN Flood attack (i.e. attack 3) on the network hosts at different times, and obtains the IDS alarm log to the node unit time. The statistics of the attacks received within the attack are calculated based on the threat degree and the number of attacks. Details of the number of cyber attacks are shown in Fig. 2

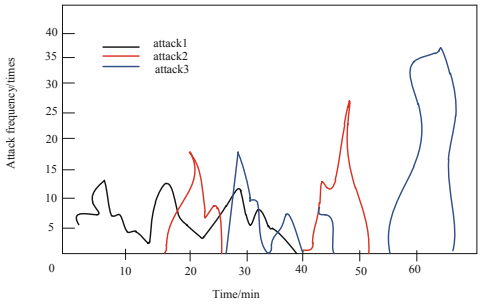

Fig. 2. Attack intensity change chart

Figure 2 shows the curve of the number of network attacks over time during the simulation experiment. The abscissa indicates the attack time in minutes and the ordinate indicates the number of attacks per minute. Attack 1 represents the attacker's scanning attack. The attack time range is between 0 and 40 min, and the threat level is relatively low. Attack 2 represents the attacker's buffer overflow attack between 15 to 30 min and 40 to 55 min respectively. Attacking two different hosts has the highest security threat to the system. Attack 3 indicates that the attacker uses TCP-SYN Flood attack and attacks the HTTP server between 25 to 40 min, 40 to 45 min, and 55 to 65 min, and is a medium-threatening attack.

The experimental data will now be used for the prediction of the network security situation.

3.2 Experimental Results and Analysis

Using the system designed in this paper and the traditional system respectively to perform network security situation composite prediction, the prediction results are shown in Fig. 3.

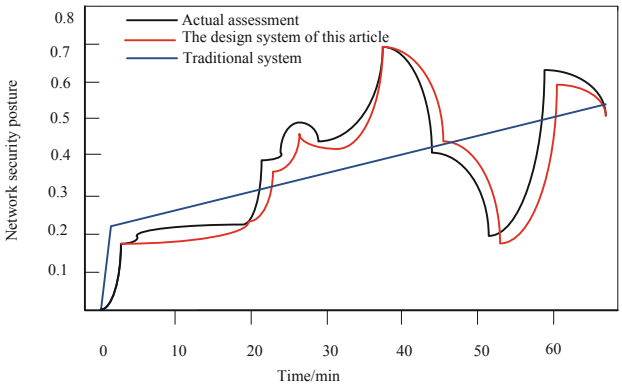

Fig. 3. Network security situation composite forecast

From Fig. 3, it can be seen that the system was scanned in the first 15 min, and the threat value increased, but the system was in a state of detection because of the low threat of the scan attack, and the risk was not very high. In 15 to 35 min, the system is moderately attacked, and the security assessment value is higher, indicating that the network is under attack. In the period of 40 to 45 min, the system is subjected to high intensity buffer overflow attack, the security evaluation value increases rapidly, the whole network is in extremely dangerous state, and the network is required to be safely protected. No attacks were detected between 45 and 50 min, but the danger faced by the network was not eliminated, and the security situation values gradually declined. From 55 min, the network was subjected to a high-intensity TCP-SYN Flood attack, and the security evaluation value increased again.

And from Fig. 3, it can be seen that the prediction error of the traditional prediction system is large, which can only reflect the general trend of the situation, but the prediction error of the design and prediction system is very small, which basically reflects the general trend of the network security situation, that is, the prediction accuracy of the design system is higher than the traditional system, and it is more suitable for the network security state. Potential prediction

This paper compares the prediction error of the system with the traditional system, as shown in Table 1.

Table 1. Comparison of the error of prediction results.

Name	Mean absolute error	Root mean square error
Traditional system	3.505	18.1466
The system of this article	0.9718	1.2823

From the data of Table 1, we can see that compared with the traditional system, the prediction results of this design system have less error and higher accuracy. Compared with the traditional system, the accuracy of the system prediction results is improved by about 65%. The prediction results of this system have higher practical value. The above two experimental results show that the prediction accuracy of the system is higher than that of the traditional system, and the accuracy is increased by about 65%.

4 Concluding Remarks

In this paper, the large-scale network security situation prediction system is optimized, and the SVM algorithm is used to improve the accuracy of prediction results. The test data show that the accuracy of the system designed in this paper is about 65% higher than that of the traditional system, and it has high effectiveness. It is hoped that this study can provide useful help for large-scale network security situation prediction.

References

1. Zhang, X., Pang, T.: An approach to real-time network security situation prediction. Sci. Technol. Vis. **12**(3), 289 (2017)
2. Xin, J., Zhang, C., Li, W., et al.: A power communication network security situation prediction method and system: CN107124394A, 15(04), 110–111 (2017)
3. Man, L.I.: Research on network security situation awareness and prediction of power monitoring system. Inf. Secur. Technol. **8**(10), 60–62 (2017)
4. Wei, H.: Discussion on the methods of power information network security situation assessment and prediction. Netw. Secur. Technol. Appl. **12**(1), 120–121 (2017)
5. Zhang, J., Zhang, B., Shen, Q.: Information system security situation assessment based on data mining. Mod. Electron. **40**(21), 77–79 (2017)
6. Chen, Y., Zhu, B., Tian, H., et al.: A network security situation analysis method based on neural network and big data and its system: CN106302522A 17(22), 1200–1203 (2017)
7. Yan, J., Weihong, H., Wei, W.: Design and implementation of large scale network security situation analysis system YHSAS. Inf. Technol. Netw. Secur. **4**(1), 222–223 (2018)
8. Zhaojun, G., Ruili, W., Shuaiqing, W.: Research on real-time forecast of security posture of information system based on improved Grey-Markov chain. Comput. Appl. Softw. **34**(2), 272–279 (2017)
9. Duanli, W.: Network to detect abnormal data in the database optimization simulation. Comput. Simul. **34**(5), 410–413 (2017)
10. Zeng, B., Zhong, P.: Simulation study on network security situation forecast. Comput. Simul. **29**(5), 170–173 (2012)