Simulation of Computer Network Information Security Assessment Model Based on Data Mining

Yanjia Liu
Institute of Computing Technology
Chinese Academy of Science
liuyanjia@ict.ac.cn

Baojian Hua
School of Software Engineering
University of Science and Technology of China
bjhua@ustc.edu.cn

***Abstract*—With the continuous growth of science and technology, network technology has been widely used, which has brought great convenience to people's life and work. However, the rapid improvement of network technology has also given network viruses an opportunity, which has caused harm to the information security of computer users. In order to improve the adaptability to the network environment, it is more effective to use data mining (DM) to prevent and control viruses. Through the use of this technology, the control effect of network security management can be improved, and the effective support for the later network security development can be realized. This paper will discuss the related modes of DM, study the practical application of DM to computer network virus defense, and build a computer network information security assessment model based on DM. The simulation results show that the proposed model can achieve a recall rate of 94.8% and an accuracy rate of 95.5%, both of which are better than the random forest model. DM has been effectively applied in the stage of protecting computer network virus, which greatly improves the security of computer network system, and then ensures the stability of computer system operation.**

***Index Terms*—Data mining; Computer network; Information security; Virus defense**

I. Introduction

With the rapid improvement and growth of computer technology and internet technology, the degree of social informatization has been continuously improved, forming a global integrated information society. As a result, people's dependence on computer networks has increased rapidly, and the importance of computer network information security has been paid more and more attention [1]. The function of modern network information system is becoming more and more complex, and the network system is becoming more and more powerful [2]. Because of the interconnection, extensibility and complexity of network and information system, their development and application will be threatened by network hackers, Trojans, viruses, malicious codes, physical failures, man-made destruction and other aspects, which has brought great inconvenience and economic losses to the country and people [3].

As a new network monitoring method, network security situation awareness refers to the perception, acquisition, understanding, assessment, display and prediction of future trends of the elements involved in the security situation in the network environment [4]. In order to improve the adaptability to the network environment, it is more effective to use DM to prevent and control viruses. Through the use of this technology, the control effect of network security management can be improved, and the effective support for the later network security development can be realized.

Because the economy of modern society is changing with the changes of the times, in the stage of continuous growth of economy and information, computer network technology is also undergoing tremendous changes, which has been applied to people's lives and brought convenience to people [5]. Computer network viruses often endanger computer security, but also cause corresponding difficulties to people, which will make people's files and all kinds of data stored on the computer infringed [6]. A large quantity of viruses emerge, because these viruses are various, complex and changeable, and their existence also gradually reduces the safety factor of computer networks [7]. In case the virus enters the network system, the user's system information and data will be stolen, which will cause the user a lot of losses, and the hardware system may be paralyzed most seriously [8]. If a virus invades the computer, it will not only harm the privacy, account information and confidential files of the system users, but also threaten the hardware facilities of the computer, even paralyze it [9]. Therefore, the application of DM in computer network virus defense provides a guarantee for the security and stability of computer network system. This article will discuss the related modes of DM, study the practical application of DM to computer network virus defense, and build a computer network information security assessment model based on DM.

II. Methodology

A. Characteristics of network virus

Computer network information security refers to a comprehensive technology that the hardware, software or other system data in the computer network system are protected from being destroyed, changed and leaked due to malicious plug-ins or software, so as to ensure the normal operation of the computer. Because the production of viruses is relatively simple, and virus programs are easy to modify and write, it makes the derivation of viruses very easy, leading to a wide variety of viruses and uncertainty. With the growth of computer and Internet technology and the continuous expansion of information resources, computer systems have gradually

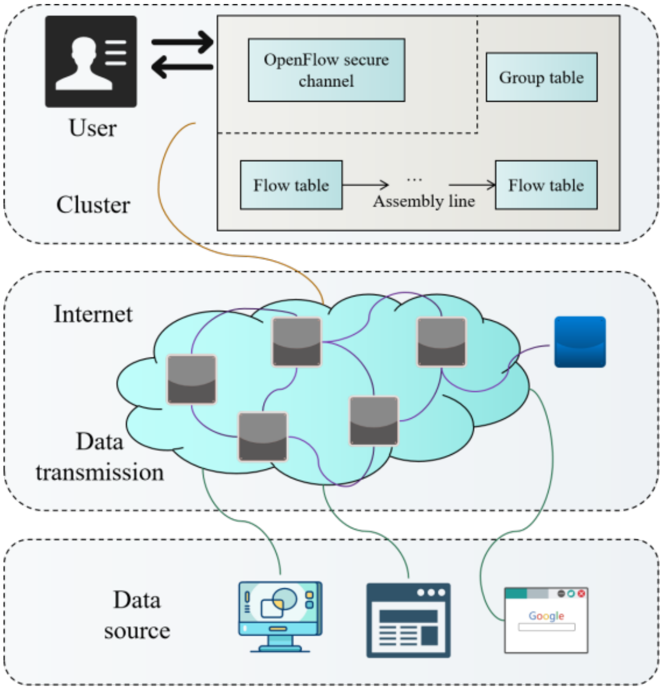

Fig. 1: Deployment of computer network information processing platform.

become complex, and the information security of computer networks has gradually been threatened by illegal invaders [10].

Viruses spread mainly through the Internet, which can spread through phishing websites, e-mails and loopholes in computer network systems, and achieve the purpose of invading computers through these channels. A virus refers to inserting a set of computer instructions into a computer program, which can destroy the functions or internal data of the computer and replicate itself. Trojan horse refers to a malicious code with special functions hidden in normal programs, which can destroy and delete files [11]. The comparison of traditional computer network virus transmission modes and the harm caused by it are universal, but with the rapid growth of network technology, network viruses are constantly improving, and some highly targeted viruses have appeared. That is to say, the production of network viruses is more purposeful, and the production and spread of viruses are more commercialized, with the main purpose of illegally seeking benefits.

Generally speaking, network security is mainly to ensure that users will not be intercepted and viewed by external intruders when transmitting data. The fundamental significance of network security lies in improving the safety factor of network system, avoiding computers from being illegally controlled or invaded by viruses, and ensuring the safety and integrity of information resource database in network system. The deployment schematic of the computer network information processing platform is shown in Fig. 1.

Network hackers implant viruses through network technology to attack and invade users' computers, destroy users' computer network systems, and steal users' information, which not only threatens users' privacy security, but also damages users' economy. Once the computer is invaded by network virus, not only the user information will be lost, but also the computer will not run normally or even be paralyzed. Before carrying out computer network information security protection in network technology, the staff need to ensure the physical security of the computer room, such as paying attention to the location of fire exits and entrances of the computer room and the registration of personnel entering the computer room, etc., so as to create a safe network environment by means of multiple physical protection.

B. *Computer network information security assessment model*

Compared with other viruses, virus programs based on computer code are somewhat complicated, not only in terms of production, but also in similarities with computer software programs [12]. For network virus defense, DM can play an effective role in virus defense, and can also ensure the security of computer networks [13]. In DM, its data collection function can accurately collect the data of network virus transmission routes and the data composed of network viruses. In the stage of computer network virus invading the computer, the invasion is mainly realized in the form of code. In the stage of virus damaging the computer, the computer program needs to be supported. In the operation of computer network, it is often because of external environment or human factors that the network is unstable and hidden dangers are buried.

With the gradual popularization of computer networks, security risk factors will increase proportionally with the increase of the quantity of users. Computer information security protection strategy is a complete system, which can be formulated and implemented through reasonable research. The hierarchical model of computer network information security is shown in Fig. 2.

Data encryption refers to the process in which the transmitter converts the computer network data into meaningless ciphertext through encryption keys and encryption functions, and the receiver restores the ciphertext to the original computer network data. It can ensure that the computer network data will not be viewed and copied by illegal means during transmission, thus ensuring the security of the computer network data to a certain extent. Let the probability distribution of random variable set

$$X = \{X_1, \ldots, X_n\}$$

be

$$P = \{P_1, \ldots, P_n\}.$$

If all variables are $\{0,1\}$, $2^n - 1$ parameters are needed to determine the joint distribution. And through Bayesian formula, the joint distribution can be written as:

$$\begin{aligned} &P(X_1, X_2, \ldots, X_n) \\ &= P(X_1)P(X_2|X_1)\ldots P(X_n|X_1, X_2, \ldots, X_{n-1}) \\ &= \prod_{i-1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1}) \end{aligned}$$

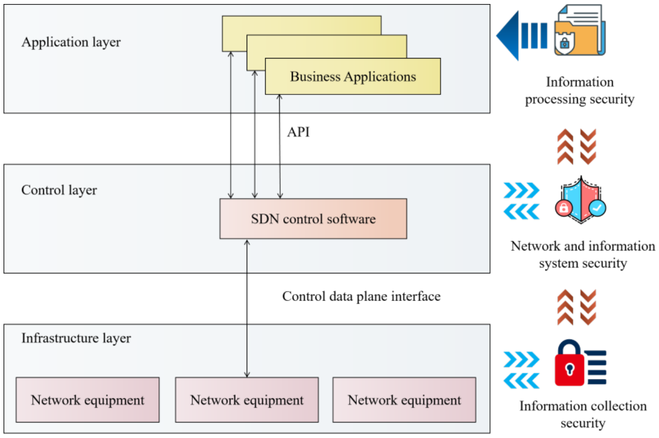

Fig. 2: Hierarchical model of computer network information security.

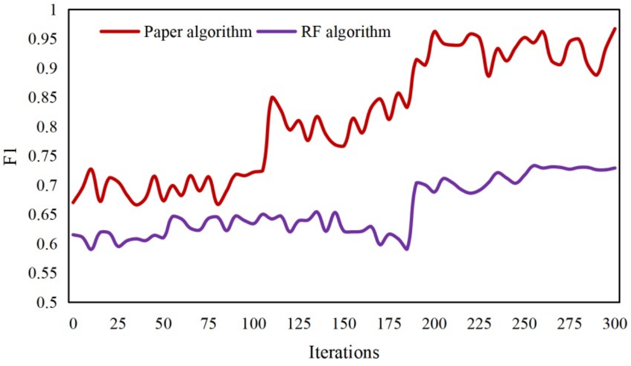

Fig. 3: Comparison of F1 value results.

For $\forall X_i \in X$, if $\pi(X_i) \subseteq \{X_1, X_2, \ldots, X_{i-1}\}$ exists, the conditions of X_i assimilation $\{X_1, X_2, \ldots, X_{i-1}\} / \pi(X_i)$ are independent when $\pi(X_i)$ is given, and the above formula can be changed to:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P\left(X_i | \pi\left(X_i\right)\right)$$

DM has the function of data processing. The application of data processing function in virus defense needs to analyze and filter the data first, then collect the virus code, and then convert its format. The biggest difference from passive attacks is that hackers will take detailed measures when they carry out attacks, which can tamper with data information or refuse related services. Active attacks cannot take preventive measures, and attacks are easy to be found by users. For active attacks, intrusion detection systems or firewalls need to be set up to improve network security from the perspective of protection. In practice, encryption can be carried out from the source and the propagation. If the transmitter wants to ensure the security of information transmission, it needs to apply encryption function and key conversion information to turn it into meaningless ciphertext. Assume that the transmission delay of the i smart grid task is T_{ui}. If it is assumed that the uplink and downlink transmission rates of any T_{ui} service access to the site are fixed at r_i, and when the quantity of tasks S_i transmitted to the internal server of MEC server is D_{ui}, the transmission delay can be expressed as:

$$T_{ui} = \frac{D_{ui}}{r_i}.$$

If the task S_i is executed at the mobile terminal, the total system delay from sending the uninstall request to the completion of the task execution can be expressed as:

$$TC_i = T_{ti} + tb_i + \min(P_{ci,j} + t_i)$$

T_{ti} represents the transmission delay, tb_i represents the delay waiting caused by insufficient bandwidth, $P_{ci,j}$ represents the delay caused by queuing at MEC, and t_i represents the delay caused by MEC server when executing tasks. If task S_i executes related commands on MEC server, its execution delay can be expressed as:

$$TC_{i,j} = T_{ti} + tb_i + P_{ci,j} + t_i.$$

Computer programs that have adverse effects on unprotected computers or data planting, repeatedly copy these harmful programs, and affect the normal operation of computer programs. With the growth of computer technology, computer viruses are also developing, which requires the constant updating of computer virus killing technology to ensure the security of computer network information.

III. Results Analysis and Discussion

There are many risks in computer network information, especially malicious attacks by hackers, tampering with basic programs by using system vulnerabilities, and even implanting viruses that destroy the system, resulting in hidden dangers in network information security [14]. In addition, there are risks of the computer system itself, such as unreasonable configuration settings, failure to regularly check the virus and update the system, which leads to the invasion and spread of the virus, resulting in damage to the motherboard, abnormal traffic, loss or modification of files, etc. In order to ensure that the application of assessment algorithm plays a benign role in network security management, the system needs to analyze the network system according to the standardized assessment mode to judge whether it is invaded.

Among all kinds of threats faced by computers, remote intrusion has the characteristics of high degree of danger and high concealment among network information security threats. When network intrusion occurs, it is difficult for ordinary network users to realize that they are stealing information. Faced with this situation, computer users can enable the security audit function of computer system logs to protect against remote intrusion threats. When computer network information transmits data information, network data will be destroyed if there is no relevant protection measures. Draw the random forest model and the F1 value of this model into a line chart. As shown in Fig. 3.

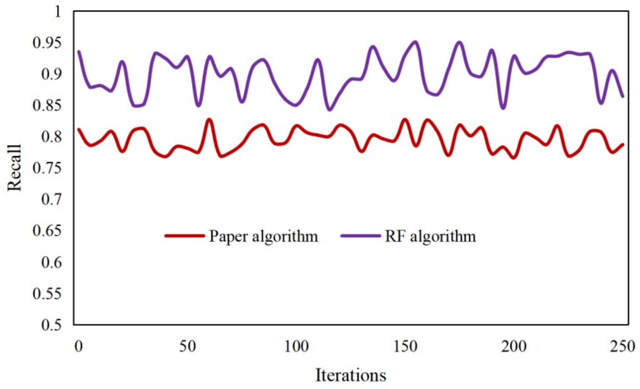

Fig. 4: Comparison of recall rates of different models.

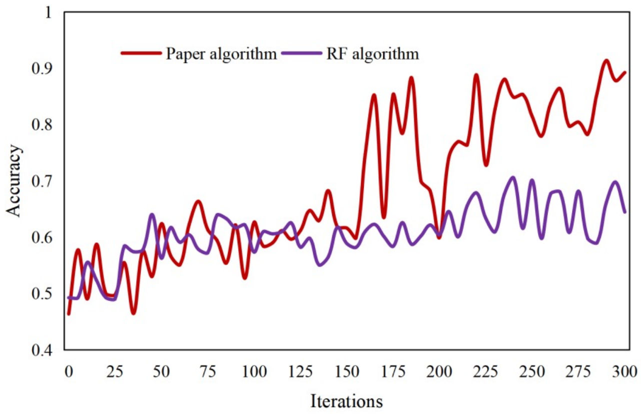

Fig. 5: Comparison of accuracy of different models.

While identifying the vulnerability, the evaluator should confirm the effectiveness of the safety measures taken. The validation of security measures should evaluate its effectiveness, that is, whether it really reduces the vulnerability of the system and enhances the ability of the system to resist risks. Effective safety measures should be maintained to avoid unnecessary waste. The safety measures determined to be inappropriate should be verified whether they are cancelled or revised, or replaced by more suitable safety measures.

The security audit function of computer system log can help users to inquire whether the computer network is remotely controlled or invaded by criminals. According to this function, computer users can also close the related network services and related computer ports that are shared by default, which can prevent the computer from being stolen by remote intrusion. The comparison results of recall rate and accuracy are shown in Fig. 4 and Fig. 5.

The results show that this model can get a recall rate of 94.8% and an accuracy rate of 95.5%, both of which are better than the random forest model. After the above contents are completed, it is necessary to analyze the correlation coefficient by using the target vector and the simulation output to obtain the slope and intercept values in the best state respectively, so as to determine whether this method is consistent with the nonlinear characteristics of network security. The input in the computer network information security assessment system comes from different data sources, including the security devices deployed in the network and the data reported by different users, etc. The diversity of data sources leads to the diversity of the collected network security event formats. Therefore, the first step in computer network information security assessment is to preprocess and integrate the data, and store the collected data into the corresponding database through some expert knowledge provided by the upper application.

IV. Conclusion

With the rapid improvement and growth of computer technology and internet technology, the degree of social informatization has been continuously improved, forming a global integrated information society. With the continuous progress of computer network technology, computer network information security protection strategies and assessment algorithms need to be continuously improved. This article will discuss the related modes of DM, study the practical application of DM to computer network virus defense, and build a computer network information security assessment model based on DM. Different from the traditional sampling method, this article innovatively puts forward the selection strategy of representative weight, which can greatly reduce the scale of candidate weight set and improve the efficiency of the algorithm. On the basis of representative weights, a deletion strategy of candidate object set is proposed to further improve the execution efficiency of the algorithm by deleting those objects that cannot appear in the result set as early as possible. The results show that the proposed computer network information security assessment model can get a recall rate of 94.8% and an accuracy rate of 95.5%, both of which are better than the random forest model. DM has been effectively applied in the stage of protecting computer network virus, which greatly improves the security of computer network system, and then ensures the stability of computer system operation.

References

[1] Gao J. A support vector machine model for network security technology. Boletin Tecnico/Technical Bulletin, vol. 55, no. 12, pp. 564-568, 2017.

[2] Yi Q. Security and Wireless Communication Networks. IEEE Wireless Communications, vol. 27, no. 3, pp. 4-5, 2020.

[3] Wen L. Security Evaluation of Computer Network Based on Hierarchy. International Journal of Network Security, vol. 21, no. 5, pp. 735-740, 2019.

[4] Zuo C. Defense of Computer Network Viruses Based on Data Mining Technology. International Journal of Network Security, vol. 20, no. 4, pp. 805-810, 2018.

[5] Ni Z, Li Q, Liu G. Game-Model-Based Network Security Risk Control. Computer, vol. 51, no. 4, pp. 28-38, 2018.

[6] Qian Y. 5G Wireless Communication Networks: Challenges in Security and Privacy. IEEE Wireless Communications, vol. 27, no. 4, pp. 2-3, 2020.

[7] Yang L. An optimized prefix span algorithm for network security and intrusion detection system simulation. Boletin Tecnico/Technical Bulletin, vol. 55, no. 11, pp. 277-284, 2017.

[8] Shi K L. Research on the Network Information Security Evaluation Model and Algorithm Based on Grey Relational Clustering Analysis. Revista de la Facultad de Ingenieria, vol. 14, no. 1, pp. 69-73, 2017.

[9] Chen J, Zhao F, Xing H. Research on Security of Mobile Communication Information Transmission Based on Heterogeneous Network. International Journal of Network Security, vol. 22, no. 1, pp. 145-149, 2020.

[10] Su Y, Han G, Fu X, et al. The Physical Layer Security Experiments of Cooperative Communication System with Different Relay Behaviors. Sensors, vol. 17, no. 4, pp. 781, 2017.

[11] Wang S, Zhu L. A markov game model of network security in information system based on copula theory. Boletin Tecnico/Technical Bulletin, vol. 55, no. 12, pp. 227-232, 2017.

[12] Li L. Integration of information security and network data mining technology in the era of big data. Acta Technica CSAV (Ceskoslovensk Akademie Ved), vol. 62, no. 1, pp. 157-165, 2017.

[13] Zhang D, Hu Y, Cao G, et al. Dataflow Feature Analysis for Industrial Networks Communication Security. Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University, vol. 38, no. 1, pp. 199-208, 2020.

[14] Shin J, You I, Seo J T. Investment Priority Analysis of ICS Information Security Resources in Smart Mobile IoT Network Environment Using the Analytic Hierarchy Process. Mobile Information Systems, vol. 2020, no. 3, pp. 1-11, 2020.