EasyChair Preprint

№ 6661

Research and Application of Terminal Vulnerability Mining System of Electric Power Company Based on Artificial Intelligence

Gu Yangqing, Yang Yu and Li Heting

September 23, 2021

Research and Application of Terminal Vulnerability Mining System of Electric Power Company Based on Artificial Intelligence

Yangqing Gu[1], Yu Yang[2], Heting Li[3]

[1]Science and technology Internet Department, Suzhou power supply branch of State Grid Jiangsu Electric Power Co., Ltd, Suzhou Jiangsu, China

[2]Science and technology Internet Department, Suzhou power supply branch of State Grid Jiangsu Electric Power Co., Ltd, Suzhou Jiangsu, China

[3]Science and technology Internet Department, Suzhou power supply branch of State Grid Jiangsu Electric Power Co., Ltd, Suzhou Jiangsu, China

Abstract: In the era of Energy Internet, prefecture and municipal power supply companies are faced with a large number of underlying terminal equipment. Traditional manual detection of vulnerabilities has the disadvantages of low batch processing efficiency and unstable vulnerability detection rate. This article combines vulnerability mining and artificial intelligence, using bidirectional LSTM network. Feature learning is used as the core algorithm of the automatic vulnerability mining module, and has been applied in practice in the world-class urban power distribution network demonstration project in the ancient city of Suzhou. The application results show that the terminal vulnerability mining system based on artificial intelligence can greatly improve the automatic discovery ability and analysis efficiency of terminal equipment high-risk vulnerabilities, and can effectively improve the defects of traditional methods that cannot be batched and modularized.

Keywords: Energy Internet; Artificial Intelligence; Vulnerability Mining; Bidirectional LSTM

0 Introduction

With the continuous development of the construction of the Energy Internet in Suzhou power supply company, the types of terminal business are increasingly diversified, such as charging piles for new energy vehicles, power monitoring, relay routing, gateways, sensors, flow meters, substations, and transformers. This makes terminal security management and control increasingly complex; secondly, terminal device types are diverse, and zombie devices that are not effectively managed will gradually increase, and the possibility of network attacks infiltrating the core platform from edge devices is increasing. It is not difficult to judge that the traditional safety management strategy and protection methods of Suzhou power supply company will not be able to effectively meet the construction needs under the background of the Energy Internet[1].

Security vulnerabilities are the main way for systems to encounter malicious attacks. According to existing statistics, the information system of Suzhou power supply company has blocked more than 3,000 application vulnerability attacks since its deployment. Among them, in the first half of 2020, Suzhou power supply company conducted a special inspection and governance of the internal network and general terminal network security. A total of 932 security vulnerabilities were discovered and corrected during this special security inspection. In addition, with the rapid increase of software and hardware platforms in the construction of the Energy Internet, and the massive growth of terminal devices, it is vital to discover and prevent application vulnerabilities[2]. At present, the traditional vulnerability mining technology mainly has defects such as limited code inspection speed, unable to realize automatic and batch detection, and the coupling degree of code inspection is high. The whole process from vulnerability discovery, verification to utilization mainly depends on the experience of security engineers. Vulnerability mining can not realize modularization, which is one of the factors restricting the speed of code inspection; traditional tools also have shortcomings in the accuracy and

false positive rate of code detection. At present, most mature tools are based on fixed pattern matching, unable to deeply understand the code semantics, and it is difficult to achieve intelligent vulnerability detection.

At present, the world is currently at the intersection of the energy revolution and the digital revolution. Artificial intelligence is a strategic technology leading this round of revolution[3]. In the field of network security, security defense capabilities lag behind the development of attack technology to a certain extent. Judging from the current global power network security incidents, security threats include hacker intrusion, bypass control, integrity destruction, unauthorized operation, unintentional or deliberate behavior, interception and tampering, illegal users, information leakage, network deception, identity disguise, denial of service attacks, and eavesdropping. Artificial intelligence technology can be effectively integrated with other technologies to form more advanced technologies. For example, artificial detection, response and repair measures are transformed into automatic repair mechanisms, which brings huge development opportunities to network security. Therefore, the application of artificial intelligence technology to the terminal vulnerability mining of power companies will surely improve the automation and intelligence of network security protection[4-5].

Therefore, this article uses artificial intelligence technology to optimize the existing network security management. It mainly studies the application scenarios of artificial intelligence technology in the terminal vulnerability mining of Suzhou power supply company. First, it explains the mechanism of terminal vulnerability mining and analyzes its system composition; the above technical solutions are applied to the world-class urban power distribution network demonstration project in the ancient city of Suzhou. The application results show that: compared with traditional vulnerability mining methods, artificial intelligence-based vulnerability mining methods have the advantages of intelligence, high efficiency and support for batch processing, and can solve the outstanding contradiction between the mining accuracy and mining efficiency of the vulnerability detection method and the actual demand.

1 Mechanism of terminal vulnerability mining

The structure of the vulnerability mining system is mainly divided into five modules: vulnerability detection cluster deployment, vulnerability scanning unit structure design, vulnerability scanning matching library selection, data storage module deployment, and artificial intelligence algorithm module. The specific analysis is as follows:

(1) Vulnerability detection cluster deployment

At present, the "Rise Eye" system of Suzhou power supply company adopts a single server mode, as shown in Figure 1, the vulnerability scanning speed is relatively low. Taking into account the characteristics of the large number of Energy Internet devices and the large number of scans, and the dynamic changes of some mobile devices such as IP; and the deployment of the vulnerability scanning module has little impact on the performance of the entire network, this system adopts the method of scanning clusters, as shown in the Figure 2, multiple scanning nodes can be deployed at the same time to perform scanning simultaneously, and then unified storage processing. The cluster uses a master node + multiple child nodes to deploy: the master node is mainly used to process external input scanning tasks; the child nodes are used for the specific scanning process. When the master node receives the scan task, it splits the current task into several subtasks according to the scope and quantity of the scan required by the current task, and distributes the subtasks to each child node through middleware; each child node receives the task after the information, the scanning process is started, and the real-time scanning data is sent to the data storage module.

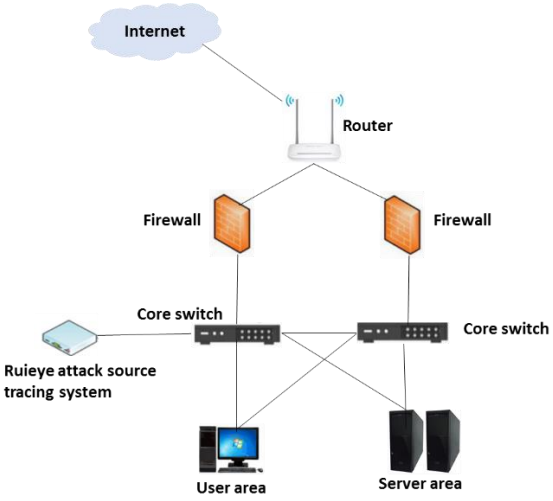

Figure 1 Existing safety protection network topology of Suzhou power supply company

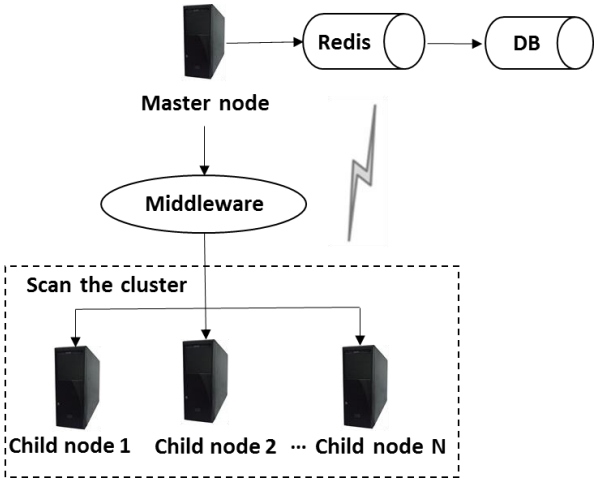

Figure 2 Optimized cluster-based network vulnerability scanning structure

(2)Vulnerability scanning unit structure design

Each scanning node is a scanning unit. Each scanning unit is essentially an independent scanning system with completing functions of the entire scanning process. When scanning a single object, it is generally divided into the following steps: host detection, port discovery, fingerprint matching, device information extraction, vulnerability matching, and vulnerability verification. For tasks with multiple scanning objects, the thread pool method is adopted to improve the response speed of the system. When a task arrives, it can be executed immediately without waiting for the creation of a new thread by reusing existing threads; at the same time, multiple objects are scanned to achieve the effect of concurrent scanning. For the most critical fingerprint matching process, the dynamic configuration method is adopted. Only the fingerprint detection template needs to be generated in the specified format, and the template description information is added, then it can be integrated into the system, which is convenient for subsequent addition and maintenance. Based on the device fingerprint identification, the identity identification of the scanned device can be realized.

(3)Vulnerability scan matching library selection

The vulnerability library integrates three official vulnerability libraries: CVE (Common Vulnerabilities and Exposures), CNVD (China National Vulnerability Database) and CNNVD (China National Vulnerability Database of Information Security). CNVD and CNNVD can be associated with CVE vulnerabilities respectively. According to the relationship between the three vulnerabilities, the three vulnerabilities are integrated into one to build their own vulnerability library. At the same time, after formatting the product list associated with each vulnerability, the vulnerability matching index is built in a three-stage structure of manufacturer, model and version; the vulnerability matching service

program is designed and deployed on the main node server for centralized management. Each child node is related to the vulnerability matching service program through CVE, and obtains the vulnerability data in the matching.

(4) Data storage module deployment

Since the rule base based on artificial intelligence needs to be based on a large amount of historical data, the data storage adopts the combination of redis + mysql secondary cache and the deployment of independent storage servers. Redis is used to cache temporary data, mysql is used to store classified scan results. The system is deployed in a scanning cluster mode, the redis database is deployed on the master node to cache temporary data. After each child node scans the specified object, it does not need to interact with the master node, and directly stores the scanned data in the redis database. An independent program reads the cached data from redis, and categorizes and aggregates the data, and stores it in mysql. When the warehousing performance is not enough, multiple warehousing programs can also be opened to read the cached data and store them at the same time. The system scanning cluster design, redis cache storage design, fingerprint detection and other designs are all dynamically deployed, which can be flexibly adjusted according to on-site equipment conditions and performance, and scanning depth requirements.

(5) Core module algorithm

There are many excellent algorithms and models in the field of deep learning, including Neural Networks (NN), Recursive Neural Network (RNN), Long Short-Term Memory (LSTM), and Bidirectional Recursive Neural Network (Bidirectional Recurrent Neural Network, BRNN), etc. The dynamic instrumentation tool pin is used to extract the instruction sequence during the execution of the terminal program, and the bidirectional LSTM network is used for feature learning, and the effect is better. This is used as the core algorithm of the Suzhou power network terminal vulnerability mining system[6-8].

As shown in Figure 3, it is the basic network structure of LSTM. Each block in LSTM mainly contains one or more cyclically connected memory cells and input gates, output gates and forgetting gates. LSTM has a gate unit control mechanism, which allows the network to "forget" the past input state stored in the memory unit, and completes operations such as adding or deleting information. Let c_t denote the state of the memory unit at time t, and update c_t in the following way:

$$\text{Input}_{\_t} = \text{sigmoid}(W_{xi}^T x_t + W_{hi}^T h_{t-1} + b_i)$$

$$\text{Forget}_{\_t} = \text{sigmoid}(W_{xf}^T x_t + W_{hf}^T h_{t-1} + b_f)$$

$$\text{Output}_{\_t} = \text{sigmoid}(W_{x0}^T x_t + W_{h0}^T h_{t-1} + b_0)$$

$$\text{Update}_{\_t} = \tanh(W_{xg}^T x_t + W_{hg}^T h_{t-1} + b_g)$$

$$\text{C}_t = \tanh(\text{Forget}_{\_t} \bullet \text{C}_{t-1} + \text{Input}_{\_t} \bullet \text{Update}_{\_t})$$

$$\text{h}_t = \tanh(\text{Output}_{\_t} \bullet \tanh(C_t))$$

In the formula, sigmoid and tanh are activation functions, and x•y represents matrix multiplication. x_t represents the current input vector, w_{xi}, w_{xf}, w_{x0}, w_{xg} represent the weight matrix of the neurons of each layer connected to the input vector x_t, w_{hi}, w_{hf}, w_{h0}, w_{hg} represent the neurons of each layer connected to the previous hidden state h_{t-1}, b_i, b_f, b_o, b_c are deviation vectors. $\text{Input}_{\_t}$ represents the input gate vector, $\text{Output}_{\_t}$ represents the output gate vector, $\text{Forget}_{\_t}$ represents the forget gate vector, $\text{Update}_{\_t}$ is the state update vector, and $\text{hidden}_{\_t}$ is the output hidden state vector. C_t represents the updated unit, C_{t-1} represents the old unit, and h_t represents the output hidden state unit vector, which is the final result.

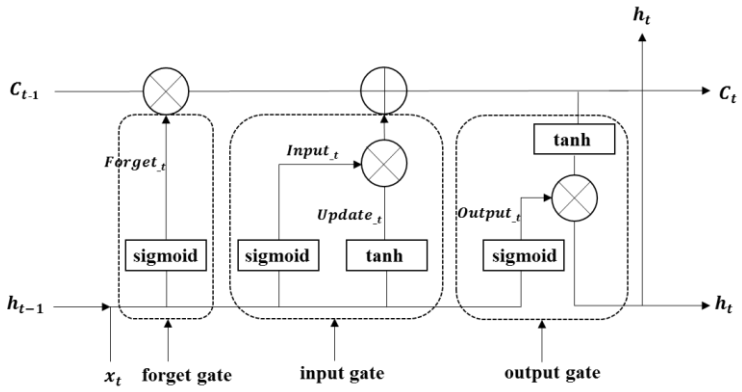

Figure 3 The structure of a bidirectional LSTM

2 Practice and improvement effect of network security artificial intelligence in the ancient city of Suzhou

According to the previously proposed terminal vulnerability mining system scheme based on artificial intelligence, the pilot deployment and application are carried out in the ancient city of Suzhou. The key implementation steps include overall system deployment map optimization, artificial intelligence technology deployment, configuration verification, data set construction, performance test and application effect analysis. From the test results, compared with the traditional methods, the technology based on artificial intelligence has excellent performance in efficiency, accuracy and dealing with unknown security threats. It is worth further promoting in Suzhou to improve the intelligent level of network security protection of Suzhou power grid.

2.1 Smart energy demonstration project in the ancient city of Suzhou

The Suzhou's ancient city demonstration project is one of the five major demonstration projects of Suzhou's smart energy and high-efficiency urban Energy Internet. It has a good foundation for power grid construction and has the superior conditions for the advancement of artificial intelligence technology. The ancient city of Suzhou is one of the characteristic model demonstrations of Suzhou's smart energy and high-efficiency urban Energy Internet demonstration construction. Taking advantage of the block characteristics of "the country's only national historical and cultural city protection demonstration zone" and the Suzhou municipal government's "central city overhead line renovation and land entry special action" as an opportunity, we will carry out the construction of a demonstration project for the Internet of Things in Suzhou's ancient city. We will develop an edge agent system for distribution grid grids, and create a new distribution network system architecture with comprehensive perception, regional autonomy, cross-regional collaboration, information integration and business collaboration, and practice the organic integration of ubiquitous power Internet of Things and strong smart grids Energy Internet construction.

The Suzhou's ancient city demonstration project is designed in accordance with the overall architecture of the ubiquitous power Internet of Things of "cloud pipe side end". The architecture is clear and is easy to implement artificial intelligence technology deployment. In the overall architecture, distributed DTUs are transmitted to the distribution automation Area I through the grid edge agent system in Area I through the MQTT protocol. The grid edge agent system in Area IV collects the electricity and non-electricity data of multiple TTUs in the grid. TTU collects various sensor data such as plug-and-play management terminals, smart gateways, access control, and cameras. The grid edge agent realizes regional autonomy (regional intelligence) based on the grid. Important production data is summarized by the grid edge agent and sent to the cloud master station (IoT management center) in a unified manner. The cloud master station (IoT management center) is responsible for interacting with the inventory system (PMS, supply and service system, etc.), and distributing strategies and related data to the grid edge agent.

2.2 Deployment of terminal vulnerability mining system based on artificial intelligence

Since the artificial intelligence-based terminal vulnerability mining system faces the new equipment that is constantly connected to the ancient city, it is deployed at the terminal layer. The deployment diagram is shown in Figure 4.

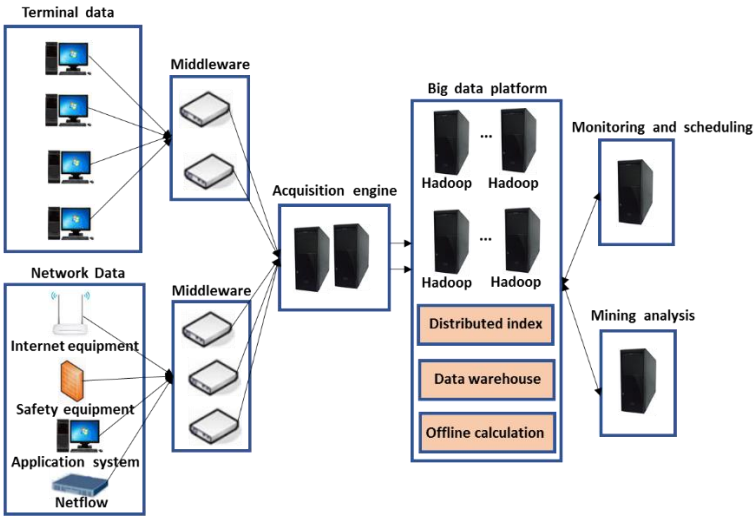

Figure 4 Schematic Diagram of Deployment of Terminal Vulnerability Mining System Based on Artificial Intelligence

In the terminal vulnerability mining system based on artificial intelligence, the specific deployment plan is as follows[9-10]:

1) Deploy a collection engine at the terminal layer of each grid in the ancient city of Suzhou. The collection engine is responsible for collecting the source code of all smart terminals;

2) Deploy a big data platform at the back end of the acquisition engine to store massive amounts of data and build a distributed index;

3) Based on the big data platform, deploy a vulnerability mining analysis module that includes artificial intelligence algorithms;

4) The vulnerability mining system preliminarily judges whether the device contains vulnerabilities through a quick scan of the newly entered source code, and judges whether the device is allowed to connect to the network;

5) After the terminal device is connected, the vulnerability mining system will continue to perform in-depth scanning and monitoring of the source code of the terminal. If a vulnerability is found, it will judge whether to continue to access the network according to the severity of the vulnerability, and issue an alarm or prompt to notify relevant management personnel.

2.3 Analysis of the role of artificial intelligence technology in the network security management of the ancient city of Suzhou

In order to verify the effectiveness of the proposed terminal vulnerability mining system based on artificial intelligence technology, according to the above deployment process, the system was installed in the ancient city network of Suzhou, data set was constructed and performance tests were performed. From this, we can get the practical effects and advantages of using artificial intelligence technology.

When testing the system, the artificial intelligence system needs to be trained before it can be put into use. Here, a network data set of the ancient city of Suzhou including the training data set and the verification data set is constructed. The basic information of the data is shown in Table 1.

Table 1 Basic information of the power network data set of Suzhou's ancient city

Type of data	Quantity
Terminal device model	25 categories
Terminal equipment system source code	3054943 lines
Terminal device operation log	66542094 articles
Terminal equipment operation record	242007 articles
Terminal interface API (program call interface) function	8030
Terminal memory usage record	523013 articles
Terminal system call frequency record	2394272 articles
TCP (Transmission Control Protocol) connection message	6756100559 articles
Network traffic log	9012517852 articles
Network delay record	6738241 articles
Attack type	38 articles
Vulnerability CVE (public vulnerability and exposure) number	372
Intrusion attack record	979 times
Network situation record	462795 articles

2.4 Analysis of application effect of terminal vulnerability mining system based on artificial intelligence

The artificial intelligence-based terminal vulnerability mining system scans the host information, account information, service information, and vulnerability information on the terminal device, and can mine the system vulnerabilities and application vulnerabilities of the scanned objects. The number of vulnerability databases is more than 2,200 articles. At the same time, you can customize the scope and strategy of mining, and customize real-time or regular vulnerabilities. The system uses artificial intelligence to build a data analysis engine, real-time mining of system vulnerabilities and protocol vulnerabilities that may exist in industrial control equipment terminals, giving full play to the advantages of parallel computing in massive data, and solving the bottleneck of data storage and analysis architecture in traditional vulnerability mining. It also adopts a cloud-to-end integrated and flat architecture to optimize the resident monitoring method of terminal equipment, with less resource occupation and high bandwidth transmission efficiency.

This paper uses the bidirectional LSTM network algorithm. Compared with the vulnerability mining system using traditional methods, the test results are shown in Table 2.

Table 2 Vulnerability mining system test comparison

Comparison item	Traditional method	Artificial intelligence-based approach
Vulnerability mining fast scanning efficiency	30 lines per second	100 lines per second
Vulnerability mining depth mining efficiency	5 lines per second	40 lines per second
Vulnerability mining fast scan accuracy rate	78 percent	90 percent
Vulnerability mining depth mining accuracy	84 percent	95 percent
Number of unknown vulnerabilities discovered	0	100
Manual participation	Need some human participation	No human involvement is required

According to the test results in Table 2, the main advantages of the terminal vulnerability mining method based on artificial intelligence are as follows:

1) Greatly improve efficiency: because the artificial intelligence algorithm adopts the parallel operation mechanism of big data, and its ability to continuously learn from samples makes the internal algorithm continuously optimized, so the operation efficiency is greatly improved;

2) Improved vulnerability accuracy: artificial intelligence algorithm can continuously enhance the learning ability, so that the judgment level and judgment knowledge of various vulnerabilities are also continuously enhanced, so as to realize more accurate judgment of vulnerabilities;

3) Ability to exploit unknown vulnerabilities: traditional methods rely on the established expert knowledge base, so the judgment of vulnerability types has limitations, while artificial intelligence algorithms can continuously evolve and learn, and even unknown vulnerabilities can be analyzed and judged through learning ability;

4) No manual participation is required: traditional methods need to manually formulate rule base, knowledge base, set parameters and auxiliary vulnerability judgment. On the premise of sufficient samples, the algorithm based on artificial intelligence does not need any human participation, which greatly reduces the workload and personnel participation.

3 Conclusion

Facing the unpredictable and increasingly severe security threats, this paper deploys and verifies the terminal vulnerability mining system in the ancient city of Suzhou. From the system operation results, it can be seen that the application of artificial intelligence technology in the project of Suzhou's ancient city area is mainly to face the access of a large number of new terminals in the power network, and endow accurate and efficient active security defense and early warning functions through artificial intelligence technology, so as to change the traditional security technology is time-consuming, labor-intensive, relying on expert experience, and lack of learning ability. The project results are of practical and practical significance, and can be further practiced and promoted within Suzhou city. Based on a higher-order vulnerability database, internal and external abnormal behaviors, and a larger training data set and verification data set, the artificial intelligence method can be optimized, and then realize the endogenous security common problems that cannot be quantitatively controlled under the

background of the Energy Internet, that is "unknown security threats", to a certain extent, converted into "knowable security issues", and provide changes for the entire power industry network security management innovative application of traditional methods.

References

[1] Chaoyang Dong, Junhua Zhao, Shuan Fu, Yusheng Xue. From Smart Grid to Energy Internet: Basic Concepts and Research Rramework[J]. Automation of Electric Power Systems, 2014, 38(15): 1-11.

[2] Shengyuan Liu, Zhenzhi Li, Jincheng Li, Yuxuan Zhao, Bo Zhang, Li Yang. Overview and Prospect of Power System Situation Awareness Technology[J]. Automation of Electric Power Systems, 2020, 44(3): 229-239.

[3] Zeqing Xiao, Haochen Hua, Junwei Cao. Application of Artificial Intelligence in Energy Internet [J]. Power Construction, 2019, 40(5): 63-70.

[4] Quanchen Zou, Tao Zhang, Runpu Wu, Jinxin Ma, Meicong Li, Chen Chen, Changyu Hou. From Automation to Intelligence: Progress of Software Vulnerability Mining Technology[J]. Journal of Tsinghua University (Science and Technology), 2018, 58(12): 45-60.

[5] Yun Li, Chenlin Huang, Zhongfeng Wang, Lu Yuan, Xiaochuan Wang. Overview of Software Vulnerability Mining Methods Based on Machine Learning [J]. Journal of Software, 2020, 31 (7): 2040-2061.

[6] Hongyu Sun, Yuan He, Jice Wang, Ying Dong, Lipeng Zhu, He Wang, Yuqing Zhang. Application of Artificial Intelligence Technology in the Field of Security Vulnerabilities[J]. Journal of Communications, 2018, 39(8): 1-17.

[7] Yun Shen, Enrico Mariconti, Pierre-Antoine Vervier, Gianluca Stringhini. Tiresias. Tiresias: Predicting Security Events Through Deep Learning[J]. In: Proc. of the 2018 ACM SIGSAC Conference on Computer and Communications Security, 2018, 592-605.

[8] Xiaojun Xu, Chang Liu, Qian Feng, Heng Yin, Le Song, Dawn Song. Neural Network-based Graph Embedding for Cross-Platform Binary Code Similarity Detection[J]. In: Proc. of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, 363-376.

[9] Xin Ning, Ping Yi. Dynamic and Static Vulnerability Mining Method Based on Deep Learning[J]. Communication Technology, 2021, 54(2): 430-436.

[10] Yuancheng Li, Rong Huang, Fenggang Lai, Yifan Mao, Lijun Cai. Open Source Software Vulnerability Detection Method Based on Deep Clustering[J]. Journal of Computer Application Research, 2020, 37(4): 1107-1110, 1114.