ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.007

A Novel Method of Fisher Discriminant Analysis for Hardware Trojan Horse Detection

Yannian Wu [1], Shaobo Zhang [1], Guohai Fan [1], Shuming Xu [1], Yiying Zhang [2] and Su Li [2+]

[1] China Gridcom Co., Ltd Beijing 102299, China

[2] College of Artificial Intelligence, Tianjin University of Science & Technology Tianjin 300457, China

Abstract. With the rise and rapid development of the network, security issues cannot be ignored. At the same time, the hardware Trojan horse to the security of the chip caused a threat. Therefore, for the detection of hardware Trojan horse in the chip, we propose a hardware Trojan horse detection method based on Fisher discriminant analysis. This method combines the discrimination ability and uses Fisher classification algorithm to extract the small difference between information, so as to realize the detection of hardware Trojan horse. However, in practice, the interference of data and the area of Trojan horse will affect the discrimination results. We conduct discriminant analysis on the number of sample chips. The analysis shows that the proposed method can effectively detect hardware Trojan horse to verify the effectiveness of the method.

Keywords: hardware trojan horse, discriminant analysis, fisher classification algorithm

1. Introduction

As the Internet becomes more and more common, the Internet has become an indispensable content for people's life and work. However, due to the openness of network information, the Internet is easy to encounter an impact, and threats are everywhere. Therefore, it is very important to secure the Internet and ensure the confidentiality and integrity of the information [1]. Meanwhile, the integrated circuits [2] are constantly developing and improving, and applying to life. However, in the process of design and manufacturing, due to the third party of dyeing fingers, it is easy for some criminals to use the designed security vulnerabilities to implant hardware Trojan horse [3] to attack, so that the chip is at risk. Hardware Trojan horse is a chip in the manufacturing process is deliberately implanted into the special module in the electronic system to destroy the normal chip. Compared to the software Trojan horse, the hardware Trojan horse is high because the logic component is small and does not affect the system efficiency when not activated, and is not easy to be checked. In addition, the design is flexible, the action mechanism is complex and the damage intensity is large. Once invaded by the hardware Trojan horse, it will pose a threat to the national interests. Therefore, how to protect the security of chips and hardware devices in the Internet is very important. It is also of practical significance for us to study the detection method of hardware Trojan horse.

In order to strengthen the understanding of the hardware Trojan horse and to create the relevant resistance systems more effectively, the researchers have done a lot of research on this, and have achieved effective results. According to the physical characteristics, the activation mechanism and the characteristic function, the hardware Trojan is classified, for improving the classification, thus promoting the development of the detection method. In addition, the researcher paid close attention to the intention of the hardware Trojan, and also put forward the corresponding detection means in the literature. For example, the hardware Trojan detection method is proposed in document [4]; document [5] uses path delay to identify some small hardware Trojan modules. Although these methods have their own advantages, they are prone to interference. Literature [6] uses genetic algorithms to generate test vectors to discover Trojans that are more difficult to

[+] Corresponding author. Tel.: 18702279203; fax: 022-60602018.
E-mail address: 1205726467@qq.com.

trigger in circuits. Document [7] distinguishes the circuit and targets the circuit structure to generate the best test vector.

However, these methods are more expensive to analyze larger circuits, and require the staff to be familiar with the internal structure of the circuit. Therefore, this paper proposes a hardware Trojan detection method based on Fisher discriminant analysis. We use the combination of hardware Trojan horses with machine learning and look for the difference between the Trojan chip and the non-Trojan chip, and then analyze the sample attributes to determine the category.

Finally, we select a straight line and complete the projection to effectively detect the hardware Trojan.

2. Guide for Hardware Trojan Horse Detection Method

2.1. Detection of Hardware Trojan Horse

With our understanding of more and more understanding of Trojan, the detection of hardware Trojan horse mainly has physical detection, functional detection, built-in self-testing technology and bypass signal analysis.

Physical detection is to let us scan the circuit one by one, and then rebuild it is according to the scanned results, and then compare the difference between the two to find the hardware Trojan in the circuit. Functional testing is to use the chip deficiencies or obstacles in the creation process to detect. Although this method can be disturbed without noise, it is difficult to detect the deficiencies and obstacles of the chip, and it can only detect the hardware Trojan that has changed the chip. Built-in self-test technology is a trusted chip producing a signature on the circuit, and a chip that contains insufficient or is implanted in a Trojan horse produces another different signature. Distinguish the Trojan chip, so as to conduct the hardware Trojan detection. Bypass signal analysis is a good detection means [8]. By collecting some power information such as temperature and electromagnetic radiation, using signal processing technology to extract the characteristic value and judge the difference, so as to determine whether there is Trojan [9] in the chip.

All these means have different attention, they have not only advantages but also flaws. In contrast, detection of bypass signal analysis is advantages. This paper combines the classification algorithm in machine learning and the hardware Trojan detection field, and selects the power consumption data of the standard chip and the Trojan chip to build the training sample data set. We use the Fisher classification algorithm to test the chip classification, so as to separate the two chips to achieve the detection effect.

2.2. Fisher Classification Algorithm

Given the training sample set, the sample is projected onto a straight line such that the projection points of similar samples are as close as possible and the heterogeneous samples are as distant as possible. When classifying the new sample, it is projected on the same line, and then the type of the new sample is determined according to the location of the projection point. This classification algorithm is also to reduce the high-dimensional problem to a one-dimensional space to solve.

Its basic idea is as follows:

Suppose there are m data in total, Given the dataset $D=\{(x_i,y_i)\}_{i=1}^{m}$, the data is projected onto the straight line w, so that the covariance of the projection points for similar samples is as small as possible. The class center distance of the heterogeneous samples was as large as possible. Thus, the target optimal solution is obtained. As shown in formula (1):

$$I=\frac{\|w^T\mu_0-w^T\mu_1\|_2^2}{w^T\sum_0 w-w^T\sum_1 w}=\frac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^T w}{w^T(\Sigma_0+\Sigma_1)w} \tag{1}$$

where μ_i, Σ_i represent the mean vector and covariance matrices of the class $i\in\{0,1\}$ sample respectively.

2.3. Detection Principle

A standard chip powered by a regulated power supply. Its power consumption can be expressed by current. As shown in formula (2):

$$I_g=I(f\cdot k)+I_{pr}+I_e \tag{2}$$

where I_g represents the total power consumption generated when the standard chip is running; $I(f \cdot k)$ is the dynamic current for operation k at frequency f; I_{pr} is the current change caused by the error; I_e is the noise generated when the chip is running.

For a hardware Trojan horse chip which exists additional malicious modules, it can produce a certain current change. As shown in formula (3):

$$I_g = I(f \cdot k) + I_{pr} + I_e + I_g(f \cdot k) \tag{3}$$

where I_g expresses as the total power consumption generated when the Trojan chip is running; $I(f \cdot k)$ is the dynamic current for operation k at frequency f. In the formula, the value of I_{pr} and I_e depends on the model of the chip and the process of data acquisition, and its impact can be weakened by taking multiple measurements of the same chip, $I(f \cdot k)$ is an important factor in judging the chip Trojan horse. According to the formula, there is a difference between the resulting power consumption, which is an important means to achieve the hardware Trojan detection by using the Fisher classification algorithm.

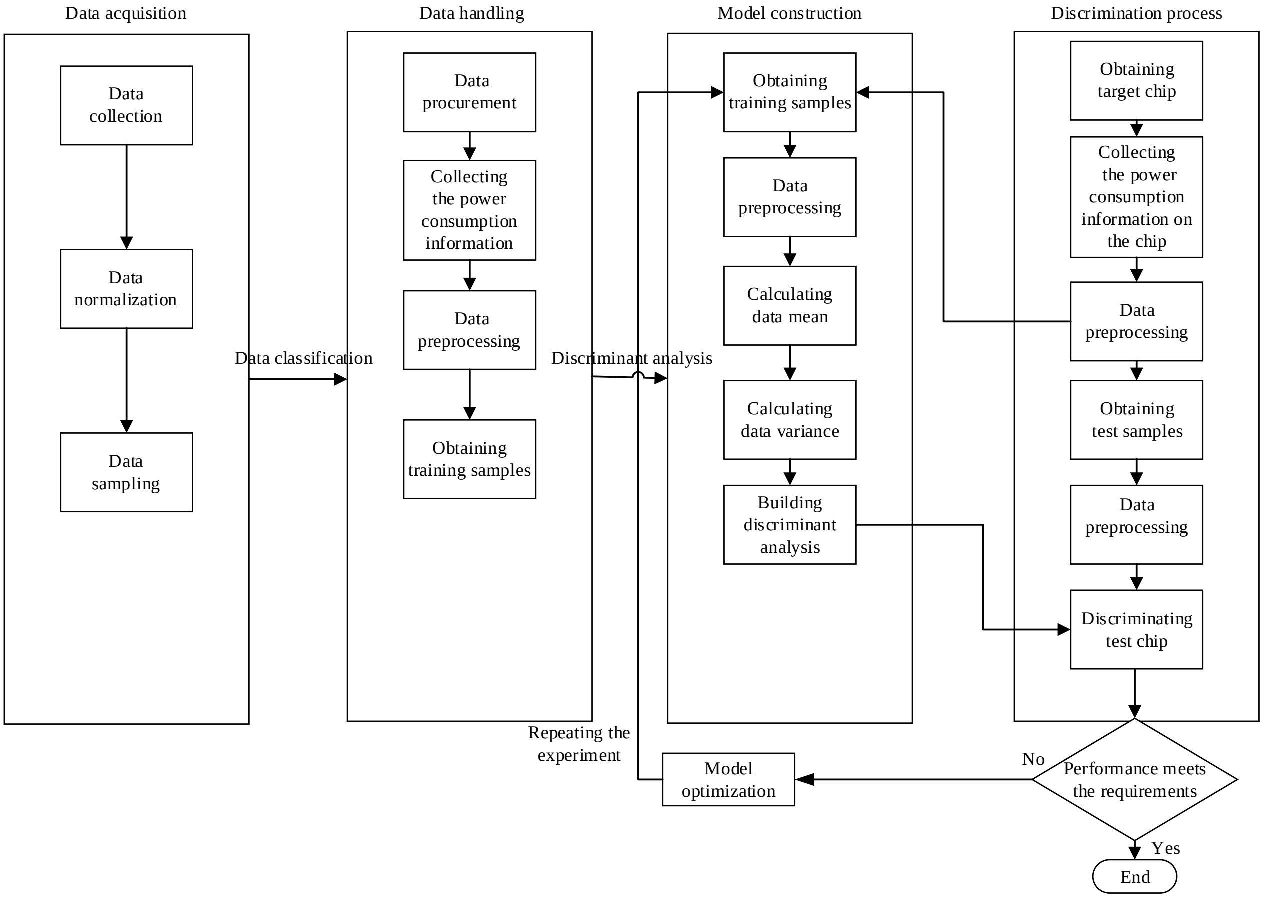


Fig. 1: Hardware Trojan horse detection model based on Fisher algorithm.

3. Algorithm Flow

According to Fig. 1, we employ the fisher classification algorithm to detect hardware Trojan horse. The process involved is mainly divided into several steps:

- Initial phase: The power consumption information of the chip to be measured is sampled, and the data is collected classified. The standard and Trojan chips are divided into training and test groups.
- Training phase: The standard and Trojan chips of the training and test groups are read separately, the mean variance of the samples is calculated, and the type is judged by using Fisher. Edit the discriminant procedure in python to judge all the data to be tested.

- Detection phase: Analyze the discrimination results to judge whether the chip contains hardware Trojan horse.
- Optimization stage: Repret the experiment and optimize to achieve the desired result.

4. Simulation Experiment

4.1. Experimental Environment

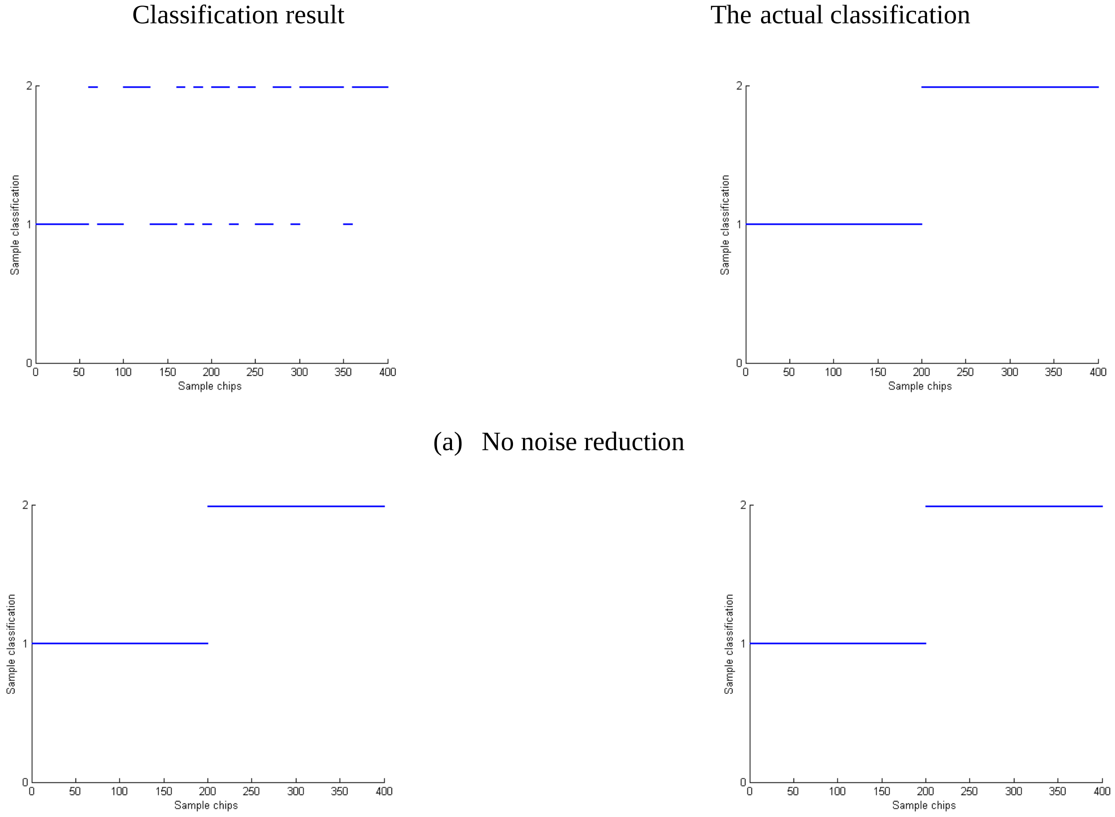
Our experiment uses the EP4CE6E22C8 chip from Altera, the type FPGA chip. The burning standard of the chip is the AES encryption algorithm program, and some of the chips are selected to join the Trojan horse module. The Teck oscilloscope DPO7104C is used to collect and display the chip power consumption information, and then received by the PC terminal. In the experiment, we use the python to edit the data. Experiments are then set up using the platform of the experiment.

In the experiment, we sample the power consumption of the measuring chip for the standard chip and the chip containing the Trojan horse. Each of the 60 groups of the standard chip power consumption data and the Trojan chip power consumption data are selected respectively as the training samples, and the 200 groups each are selected as the samples to be tested. Among them, the first 200 groups are detected as they should be the standard chip, and the last 200 groups are Trojan chips.

4.2. Experimental Results and Analysis

1) Effect of the noise on the classification results

In the experiment, interference such as noise can affect the power consumption information of the chip, so it can affect the classification result. To show the effect of noise and other interference on Fisher classification, Fig. 2 shows the results of 30 standard and Trojan chips to become training samples. The abscissa represents the chipset serial number, and the ordinate is the type serial number, where 1 and 2 represent the standard chips and Trojan chips respectively.


Classification result

The actual classification

(a) No noise reduction

(b) Each group of 30

Fig. 2: Effect of the noise on the Fisher classification.

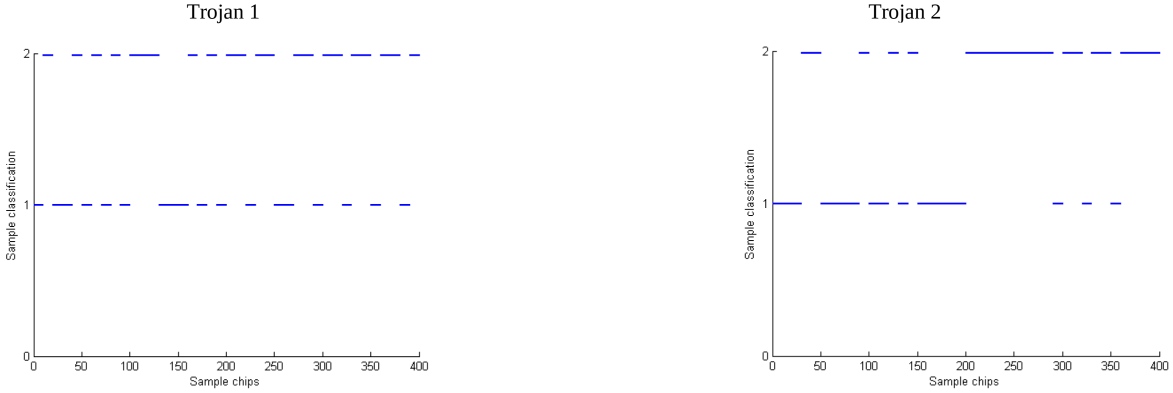
We can see from the figure that there are differences between the classification results and the actual classification without the noise reduction treatment. However, when 50 pieces of data are processed, the classification results are consistent with the actual classification. The addition of Trojan chip will cause changes in power consumption, and interference will affect the discrimination of Trojan chip. After we reduce the noise, the type of chip to be measured can be made more accurately judged. In the Table 1, we give the results of miscalculation and missed judgment in two environments.

Table 1: Noise reduction and distermination results

Whether to reduce noise	Discriminate The Analysis result(%)	
	Leakage	Erroneous judgement
Yes	0.3	0.25
No	20	7.25

From the Table 1, we can see that after reducing the noise, the proportion of missing judgment and miscalculation is significantly reduced. Therefore, the noise has a large influence on the classification results.

2) Effect of Trojan area on the experimental results


Trojan 1

Trojan 2

Fig. 3: Effect of Trojan area on Fisher classification.

Because in the analysis process, the larger the area of the Trojan horse, the more resources are consumed, and the greater the power consumption is produced. Therefore, we set up two experiments to verify the effect of its area on classification. The results of two Trojan chips with different area sizes in Fig.2, they are the discrimination analysis results of Trojan 1 and Trojan 2 occupying 3% and 5% of the original chip module respectively.

As shown in Fig. 3, the larger the Trojan area, the higher the accuracy of the Fisher discriminant analysis, the easier to distinguish. Furthermore, Table 2 presents the discrimination results obtained when planting a Trojan of two area sizes. As the results show, the Fisher discriminant analysis is more accurate for the larger area of the hardware Trojan classification under the same conditions.

The plant wood horse

Table 2: Trojan area and discrimination results

The plant wood horse	Discriminate The Analysis result (%)	
	Leakage	Erroneous judgement
Trojan 1	15	17.45
Trojan 2	7.3	5.25

3) Effect of Trojan species on the experimental results

Classification result

The actual classification

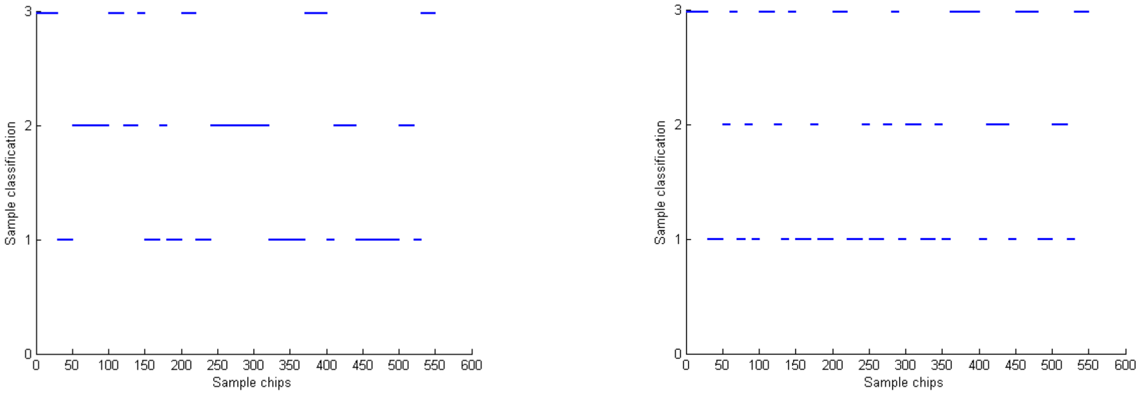


Fig. 4: Effect of Figure Trojan species on Fisher classification.

In addition to the effect of the Trojan area on the discriminant analysis, the same discriminant performance is also affected by the Trojan species. Therefore, in order to verify the influence of species, we also set up two Trojan categories to collect power consumption information and classify the Fisher classification algorithm. The result is shown in Fig. 4.

We can see from the Fig. 4 that the effect of Trojan differences on the classification results is also different. In order to more clearly see the differences between them, we compare the discriminant results with the real classification results. Based on the results of Table 3, it is also clear that the Fisher discriminant analysis correctly separates the standard chips.

Table 3: Trojan Horse Types and Discrimination Result

The plant wood horse	Discriminate The Analysis result(%)		
	Standard chip	Trojan 1	Trojan 2
Standard chip	100	0	0
Trojan 1	0	98.2	1.7
Trojan 2	0	2	97.8

5. Summary

This paper uses Fisher discriminant analysis to detect hardware Trojan. The algorithm is a classic two classification problem, it can mutual fusion with Trojan detection can achieve good results. Our paper combines Trojan horse and machine learning, and applies the internal classification algorithm to the field of hardware Trojan detection, so as to achieve the separation effect of standard chip and Trojan chip, which has realistic significance.

In the implementation process, the power consumption information of the target chip is collected, and the bypass signal is used to analyze the power consumption data. Finally, the proposed scheme is combined with experiments, and the effects of different environments on the discrimination results are analyzed in detail. The experimental results show that our scheme has certain advantages.

6. References

[1] Mona Algarni, Munirah Alkhelaiwi, Abdelrahman Karrar, et al. Internet of Things Security: A Review of Enabled Application Challenges and Solutions [J]International Journal of Advanced Computer Science and Applications Volume 12, Issue 3. 2021.

[2] Hao Yue, Xiang Shuiying, Han Genquan, et al. Recent progress of integrated circuits and optoelectronic chips[J]Science China(Information Sciences)Volume 64, Issue 10. 2021.

[3] Dong Chen, Xu Yi, Liu Ximeng; Zhang Fan; He Guorong; Chen Yuzhong.Hardware Trojans in Chips: A Survey for Detection and Prevention [J]SensorsVolume 20, Issue 18. 2020. PP 5165-5165.

[4] Wang X, Salmani H, Tehranipoor M, et al. Hardware Trojan Detection and Isolation Using Current Integration and Localized Current Analysis [C] // IEEE International Symposium on Defect and Fault Tolerance of Vlsi Systems. IEEE Computer Society, 2008:87-95.

[5] Jin Y, Makris Y. Hardware Trojan detection using path delay fingerprint [C] // IEEE International Workshop on Hardware-Oriented Security and Trust. IEEE, 2008:51-57.

[6] SAHA S,CHAKRABORTY R S,NUTHAKKI S S, et al Improved test pattern generation for hardware Trojan detection using genetic algorithm and Boolean satisfiability[C]. The 17th International Workshop on Cryptographic Hardware and Embedded systems, Saint-Malo, France, 2015: 577-596. doi: 10.1007/978-3-662-48324-4_29.

[7] XUE Mingfu, HU Aiqun and LI Guyue. Detecting hardware Trojan through heuristic partition and activity driven test pattern generation[C]. 2014 Communication Security Conference, Beijing, China, 2014:1-6. doi:10.1049/CP.2014.07 23.

[8] Shila D M, Venugopal V. Design, implementation and security analysis od Hardware Trojan Threats in FP-GA[C] // Proc of IEEE Communication and Informa-tion Systems Security Symposium. New York, NY,USA:ACM,2014:247-24.

[9] XIAO K,FORTE D,TEHRANIPOOR M. A novel built-in selfauthentication technique to prevent inserting hardw are Trojans[J].IEEE Trans Cpmput-Aided Design Integr Circuits Syst,2014,33(12):1778-1791.

[10] HEJiaji, ZHAO Yiqiang,GUO Xiaolong, et al. Hardware Trojan detection through chip-free electromagnetic mide-channel statistical analysis[J].IEEE Transactions on VeryLarge Scale Integrayion(VLSI) Systems, 2017, 25(10):2939-2948:10.1109/TVLDSI.2017.27 27 985.

[11] Vamshi Krishna Gudipati, Aayush Vetwal, Varun Kumar, et al. Detection of Trojan Horses by the analysis of system behavior and data packets[C].IEEE 2015 Long Island Systems, Applications and Technology. DOI: 10.1109/LISAT.2015.7160176.

[12] L. Yu-Feng, Z. Li-Wei, L. Jian, Q. Sheng and N. Zhi-Qiang, "Detecting Trojan horses based on system behavior using machine learning method", *Machine Learning and Cybernetics (ICMLC) 2010 International Conference*, pp. 855-860, 2010.

[13] Wen Yu; Ye Yalin; Ran Haodan, “Research on the Technology of Trojan horse detection”, 2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA). DOI: 10.1109/ICICTA49267.2019.00032.

[14] Wang Jing, "Research on Deep Detection Technology of Hidden Trojan Horse[D]". University of Electronic Science and Technology, 2010.

[15] Zhou WeiFu, Zhang Yiying, Zhang Suxiang and Yang Chengyue, "Research on Trojan Horse Virus Detection Technology Based on Characteristic Behavior Analysis[J]". Telecommunications Science, no. 11, pp. 105-109, 2014.