A Detection and Defense Technology for Information-stealing and Deceitful Trojan Viruses Based on Behavioral Features

Li Lin

(Corresponding author: Li Lin)

Department of Information Engineering, Huanghuai University
Zhumadian, Henan, 463000, China
(Email: linli2016_li@126.com)

(Received Dec. 12, 2017; revised and accepted Apr. 28, 2018)

Abstract

With the rapid development of Internet technology, computer networks play an increasingly important role in improving the living standard of people. However, it also induces events which threaten the security of Internet. Information stealing and network intrusion are always threatening network information security. In this study, the behavioral features of Trojan were analyzed, and a detection and defense system for information-stealing and deceitful Trojan virus based on behavioral features was designed and tested by simulation experiments. The simulation results demonstrated that the system has favorable accuracy and practicability, and its performance can satisfy practical applications. The system installed at a network gateway can monitor network flows and detect whether there is invasion of information-stealing and deceitful Trojan viruses, which is of great significance to the defense of network security.

Keywords: Behavioral Features; Cyber Theft; Network Information Security; Trojan Virus

1 Introduction

Network information security is an important cornerstone for the development of a modern network [6, 9]. A large amount of valuable resources and information stored in the computer network make criminals on the Internet eager to get all these valuable resources by using various hacking technologies [12]. Trojan virus is the most common technology. Information-stealing Trojan virus can acquire computer information resources by invading computers [3]. Developing Trojan virus detection and defense technology has been one of priorities today. Many scholars have made deep study [1].

Ni *et al.* [8] put forward a wavelet transform based a noise optimization method through detecting hardware Trojan viruses using a back propagation neural network technology. The experimental results suggested that the wavelet transform-based noise optimization method could eliminate high-frequency noises and make the sensitivity of detecting hardware Trojan viruses based on neural network increase from 92.2% to 99.2%. In a study of Xi [14], several defense algorithms were proposed based on the summary of the study status and key technologies of Address Resolution Protocol (ARP) and analysis of the formation mechanism of ARP bugs, and the advantages of the improved defense algorithms were introduced; finally the optimal defense algorithm was determined after testing every improved algorithm.

In this study, the behavioral features of Trojan virus were analyzed, and then an information-stealing Trojan virus detection and defense system was designed based on the behavioral features of Trojan virus. The test suggested that the system could effectively test information-stealing and deceitful Trojan virus, which has great significance to the protection of network information security.

2 Analysis on Behavioral Features of Trojan Virus

Trojan virus invades systems by concealing itself. Many behavioral features will be presented though it can hide its tracks left in the targeted system [10]. Next is the analysis of behavioral features of Trojan virus invasion at different stages.

1) Stage of Trojan virus implantation:
When invading a system, Trojan virus will deceive users to gain their trust using illegal means and then enter the target [5]. In this stage, the behavioral features of Trojan virus include attacking via bugs in software and systems, rogue programs, ports, foreign unknown e-mails, and unknown network links.

Table 1: Main behavioral features of Trojan virus and division of risk levels

Moderate	High
1. Replicate or create files in catalog; 2. Self-starting; 3. Call cmd process; 4. Bind monitor port.	1. Automatically delete files; 2. Automatically compress or decompress files; 3. Revise system time; 4. Creating file associations; 5. Do Internet Control Messages Protocol communication; 6. Conceal process; 7. Close corresponding process; 8. Automatically send e-mail outward; 9. Disguise system process or communication path; 10. Entering system process or IE process.

2) Stage of Trojan virus installation:
There are some special behavioral features when Trojan virus is installed. In this stage, the main action object of Trojan virus is itself. Therefore, it is a good stage to detect Trojan virus. The behavioral features of Trojan virus in this stage include compressing or decompressing files automatically, deleting some files automatically, automatically restarting, automatically system timing, automatically closing or opening programs, revising system configuration files, and revising the system relevance.

3) Stage of startup operation:
When Trojan virus is successfully installed in the target system, operation modes such as process and thread are needed. Generally, processes can be observed, but threads cannot [4]. Therefore, Trojan virus has to conceal itself. In this stage, the behavioral features of Trojan virus include concealing processes, calling cmd processes, closing specific processes. and transferring to other processes via a remote thread technology.

4) Stage of network communication:

a. After a system is controlled by Trojan virus, information communication is needed. Trojan virus will receive information from outside such as the new commands of the controller and transmit information to the controlling end via special communication modes [7]. It can control the target system and steal system information via those means.

b. Main behavioral features of Trojan virus and division of risk levels A corresponding database of Trojan virus behavioral features was established based on Table 1. A large number of Trojan virus features were collected and applied in the detection and defense of Trojan virus.

3 Design an Information-Stealing and Deceitful Trojan Virus Detection System

3.1 Framework of Trojan Virus Detection Model

A behavioral features-based Trojan virus detection model can classify programs based on behaviors of programs and the aforementioned database of Trojan virus behavioral features by using Bayes classifier [13]. The model designed in this study was composed of behavioral extraction, behavioral features database, a program behavioral analyzer, Trojan virus processor and user assistance.

Behavioral extraction aims at monitoring suspicious behaviors in the system and sending them to the behavioral analyzer [15]. Database of Trojan virus features included a large amount of Trojan virus behaviors, action objects of behaviors, descriptive information of behaviors and basic probability information. The program behavioral analyzer was the most important in the model. It can classify programs using Bayes classifier and perfect its classification processing ability through communicating with the features of database. User assistance could achieve human-computer interactions by sending contents which cannot be determined by the analyzer to users and present data to users.

3.2 The Main Module of the Proposed System

3.2.1 Design of Behavioral Extraction Module

To realize monitoring and capture of all program behaviors, a behavior monitoring module should be installed in the core of the operation system to monitor file system, registry, processes, storages, and communication. API-HOOK [11] was used to intercept a program system called in this study. The implementation of the behavioral monitoring system was introduced by taking the monitoring of registry as an example.

Table 2: The features - behavioral database

	Feature set database				Feedback self-adaption database	
2-7 Data sheet	Feature set sheet	Test feature sheet	Parameter sheet	Feature benchmark sheet	Classification performance sheet	System parameter sheet
Description	Storing sample feature vectors which need examination	Storing testing data of normal programs and Trojan virus	Storing API parameters	Storing various data of behavioral features after classification	Storing system classification performance description	Storing system analysis risk coefficient and classification width

Registry, the core database of an operation system, can store the setting information of systems and application programs. SSDT HOOK technology is needed in the monitoring of registry. All the operations associated to registry can be monitored under the assistance of SSDT HOOK. Registry operations in the monitoring module include creating, revising and deleting registry key items, and creating, revising and deleting registry key values. In the monitoring based on SSDT HOOK technology [16], SSDT entry should be backed up before revision, and the monitored function should be replaced by a self-defined monitoring function.

Besides API HOOK technology, static analysis of PE was used to acquire behavioral characteristic vectors. PE is a file format which can be transplanted and executed in window systems. It can store data set codes in a linear address space and analyze PE files using BK-50 scanner to acquire the features-behavioral vector of programs.

Static analysis of PE files is ineffective to programs which apply a code obfuscation technology. Therefore, API HOOK was the main body, and static analysis of PE was the assist in the design of the behavioral extraction module.

3.2.2 Features-Behavioral Database

The common behaviors of Trojan virus have been introduced in Section 2. The features- behavioral database contained various data, which could provide a reference for classification of the classifier. The behavioral features of Trojan virus include normal programs such as characteristic weight a, risk coefficient S, width of classifier *theta* and basic conditional probability. The features-behavioral database could be divided into characteristic set database and feedback self-adaption database. The classification of the behavioral characteristics database is shown in Table 2.

3.2.3 Behavioral Analysis Module

Behavioral analysis module, an important part of the Trojan virus detection model, was mainly composed of data preprocessing, program classification and feedback study. The algorithm was described as follows.

Data preprocessing included redundancy elimination, feature vector independent processing and weight calculation.

1) In redundancy processing, feature set and test feature set sheets were input. Through calculation of feature CRR and deletion of features with low relevancy, a non-redundant feature set sheet was output.

2) In feature vector independent processing, a non-redundant feature set was input. Then behavioral attributes were merged using SNCB model. Finally, an irrelevant non-redundant feature set was output.

3) In weight calculation, an irrelevant non-redundant feature set was input; then impact factors were assigned, scales were obtained, and weights were calculated. Finally, feature weights were output.

4) In classification calculation, sample feature sets, classification algorithms, unknown examples and a system parameter sheet were input; then features of examples were extracted by the classifier which has been regulated by statistical calculation of sample features; after classification, the classification results were output.

5) In classifier learning, a classification performance sheet was input, and feature set and system parameter sheets were output through incremental learning and re-learning.

3.2.4 System Response Module

After classification on running programs, corresponding operations needed to be done according to the classification results. For example, the discovered Trojan virus program needed to be cleared, and the classification conditions were fed back to users. After classification, the system stopped monitoring normal programs, and Trojan virus programs and programs which could not be determined were provided to users for processing. The system

Table 3: The detection results of the Trojan virus samples

	Huigezi Trojan virus	Dark distant control Trojan virus	Byshell Trojan virus
Number of samples	20	20	20
Correct detection number	18	19	19
False detection number	2	1	1
Detection rate	90%	95%	95%
False alarm rate	10%	5%	5%

Table 4: The testing results of thread performance

		500 Mbps	800 Mbps	1 Gbps	1.5 Gbps
Performance test on single thread	Peak CPU	87.1%	91.6%	89.3%	91.4%
	Memory usage	9.6%	11.6%	12.0%	13.1%
	Thread	1	1	1	1
	Packet loss probability	0%	0%	0.14%	1.6%
		2 Gbps	4 Gbps	6 Gbps	8 Gbps
Performance test on multi-thread	Peak CPU	114%	154%	234%	301%
	Memory usage	13.4%	15.7%	19.8%	25.6%
	Thread	2	4	6	8
	Packet loss probability	0%	0%	0.05%	0.02%

deleted Trojan virus programs once discovered and transmitted the information of Trojan virus to users. When it was difficult to determine whether a program was normal or not, the system would isolate the program and display the condition to users for determination. After determination, the system deleted or restored it according to actual conditions.

4 System Test

The detection and false alarm rates of the Trojan virus detection system were tested. The purpose of the test was to test the accuracy of the Trojan virus detection system. Rules were written according to the Trojan virus features-behavioral database.

The test process was as follows: A Trojan virus controlled the terminal was implanted into a virtual host. Then the control terminal was installed in an external computer. The categories and a number of Trojan virus samples were controlled through the computer. Detection rules were formulated according to the Trojan virus samples and features-behavioral database. Then the Trojan virus detection system was started and controlled to communicate with the host. The total number and valid number of alarms were accounted. Finally, the detection and false alarm rates of the system were calculated.

The expected detection rate was not lower than 80%, and the expected false alarm rate was not higher than 10%.

In the test, Huigezi Trojan virus, dark distant control Trojan virus, and Byshell Trojan virus were selected for testing. During testing, the three viruses were installed in virtual hosts, and an abnormal communication flow was generated. In the testing cluster, there were 100 hosts, and every host was installed with Huigezi Trojan viruses, dark distant control Trojan viruses, and Bysell Trojan viruses, containing 20 viruses for each kind of above viruses. Then the step number was detected at the cluster exit. After repeating tests following the above procedures, the results obtained were in following tables.

It is seen from Table 3 that the detection rates of those three Trojan viruses were 90%, 95%, and 95%, respectively, and the corresponding false alarm rates were 10%, 5% and 5%, respectively. The results suggest that the behavioral features-based, information-stealing, and deceitful Trojan virus detection system could detect Trojan viruses included in the features- behavioral database. The false alarm might happen because the users presented some behavioral features similar to Trojan viruses under a certain condition. The detection and false alarm rates satisfied the aforementioned expectations.

In Table 4, it is demonstrated that the packet loss gradually happened with the increase of the flow in a single-thread operation; the higher the flow, the severer the pack loss phenomenon. Thus, it could be concluded that the calculation capability of a single thread was not suitable for calculating a large flow, as an excessively large data packet could lead to overflow of bottom data, leading to a packet loss. Therefore, a single thread was only suitable for processing the data flow between 500 M and 1 G. Flow packages could increase continuously in the process of multi-thread processing. In Table 2, it is demonstrated that CPU occupancy rate and memory utilization rate are significantly improved during the multi-thread pro-

cessing; however, the packet loss gradually appeared with the increase of flow. Hence, the peak flow supported by the designed system was 10 G.

5 Conclusion

The development of Internet facilitates worldwide information sharing, but it also brings huge challenges to network security. Network development results in the spread of information stealing events. Some lawbreakers develop many information-stealing and deceitful Trojan viruses to steal information. In this study, the behavioral features of information-stealing and deceitful Trojan viruses were analyzed, then an information-stealing and deceitful Trojan virus detection system was developed based on the behavioral features, and the feasibility and performance of the system were tested. The test results demonstrated that the detection rate and false alarm rate of the system satisfied the standards, and the supportable peak flow was 10 G, which lays a reference for behavioral features-based Trojan virus detection technologies.

References

[1] A. A. Al-khatib, W. A. Hammood, "Mobile malware and defending systems: Comparison study," *International Journal of Electronics and Information Engineering*, vol. 6, no. 2, pp. 116–123, 2017.

[2] A. Anurag, "Network neutrality: Developing business model and evidence based net neutrality regulation," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 1–9, 2015.

[3] E. C. Dunn, "Trojan pig: Paradoxes of food safety regulation," *Environment & Planning A*, vol. 35, no. 8, pp. 1493–1511, 2015.

[4] F. Farahmandi, Y. Huang, P. Mishra, "Trojan localization using symbolic algebra," in *Design Automation Conference*, pp. 591–597, 2017.

[5] I. Hsiao, Y. K. Hsieh, C. F. Wang, *et al.*, "Trojan-horse mechanism in the cellular uptake of silver nanoparticles verified by direct intra- and extracellular silver speciation analysis," *Environmental Science & Technology*, vol. 49, no. 6, pp. 3813–3821, 2015.

[6] S. Islam, H. Ali, A. Habib, N. Nobi, M. Alam, and D. Hossain, "Threat minimization by design and deployment of secured networking model," *International Journal of Electronics and Information Engineering*, vol. 8, no. 2, pp. 135–144, 2018.

[7] X. W. Li, X. H. Wang, Y. Zhang, K. Y. Chen, L. Xu, "A new hardware trojan detection method based on kernel maximum margin criterion," *Acta Electronica Sinica*, vol. 45, no. 3, pp. 656–661, 2017.

[8] L. Ni, J. Li, S. Lin, *et al.*, "A method of noise optimization for Hardware Trojans detection based on BP neural network," in *IEEE International Conference on Computer and Communications*, pp. 2800–2804, 2017.

[9] E. U. Opara, O. A. Soluade, "Straddling the next cyber frontier: The empirical analysis on network security, exploits, and vulnerabilities," *International Journal of Electronics and Information Engineering*, vol. 3, no. 1, pp. 10–18, 2015.

[10] A. R. Ribeiro, S. Gemini-Piperni, R. Travassos, *et al.*, *Trojan-like Internalization of Anatase Titanium Dioxide Nanoparticles by Human Osteoblast Cells*, Scientific Reports, 6:23615, 2016.

[11] Y. Song, Y. Shen, G. Zhang, "The new INLINE hook technology combination of hard-code technology and independent code injection," in *IEEE International Conference on Software Engineering and Service Science*, pp. 521–525, 2017.

[12] A. Tayal, N. Mishra and S. Sharma, "Active monitoring & postmortem forensic analysis of network threats: A survey," *International Journal of Electronics and Information Engineering*, vol. 6, no. 1, pp. 49–59, 2017.

[13] M. Turkoglu, D. Hanbay, "Classification of the grape varieties based on leaf recognition by using SVM classifier," in *Signal Processing and Communications Applications Conference*, pp. 2674–2677, 2015.

[14] H. Xi, "Research and application of ARP protocol vulnerability attack and defense technology based on trusted network," in *Advances in Materials, Machinery, Electronics. Advances in Materials, Machinery, Electronics (AMME'17)*, pp. 090019.1–090019.7, 2017.

[15] M. Xue, A. Hu, W. Liu, *et al.*, "Detecting hardware Trojan through feature extraction in subspace domain," *Journal of Southeast University (Natural Science Edition)*, vol. 44, no. 3, pp. 457–461, 2014.

[16] Y. Zhang, H. Bi, "Anti-rootkit technology of kernel integrity detection and restoration," in *International Conference on Network Computing and Information Security*, pp. 276–278, 2011.

Biography

Li Lin who has gained the master's degree is a lecturer in the college of information engineering of Huanghuai University, Henan, China. Her interests of research include computer application, data mining and big data application. She has engaged in the teaching of computer course for 13 years. Moreover she has participated in multiple provincial-level, city-level and school-level projects, published more than 10 academic papers (more than half in Chinese core periodicals), participated in the editing of a textbook, and gained multiple honors from the school for her works.