RESEARCH ARTICLE | APRIL 18 2018

Summary of vulnerability related technologies based on machine learning FREE

Lei Zhao; Zhihao Chen; Qiong Jia

AIP Conf. Proc. 1955, 040054 (2018)
https://doi.org/10.1063/1.5033718

Summary of Vulnerability Related Technologies Based on Machine Learning

Lei Zhao*, Zhihao Chen, Qiong Jia

INSTITUTE706, THE SECOND ACADEMY, Beijing, China

*Lei Zhao2018@126.com

Abstract. As the scale of information system increases by an order of magnitude, the complexity of system software is getting higher. The vulnerability interaction from design, development and deployment to implementation stages greatly increases the risk of the entire information system being attacked successfully. Considering the limitations and lags of the existing mainstream security vulnerability detection techniques, this paper summarizes the development and current status of related technologies based on the machine learning methods applied to deal with massive and irregular data, and handling security vulnerabilities.

INTRODUCTION

With the popularization of the Internet, people enjoy the convenience brought by the information age and are also plagued by various security problems. During 2015, there were a number of worrisome events in our country [1]. This has alarmed China's Internet security status. Therefore, the purpose of this paper is to give a proper definition of security vulnerabilities, to analyze and summarize the mainstream technologies and major machine learning methods of security vulnerabilities detection, and to summarize the application of machine learning methods in the field of security. Finally, put forward the possible research direction in the future.

SECURITY VULNERABILITY DETECTION

Security Loopholes

Vulnerabilities are all elements of an information system that act on a given environment and can cause damage to parts, operations, and data in the system. In 2006, the National Institute of Standards and Technology (NIST) defined vulnerabilities in the Glossary of Key Terms for Information Security as: Threats to sources that exist in information systems, systems security processes, internal controls, or implementations The vulnerability that attacks or triggers is Vulnerability. In 2010, the definition of vulnerability given by China in "Technical Terms for Information Security Technology" (GB/T25069-2010) is that information technology, information products and information systems will meet the needs, design, implementation, configuration and operation of the Vulnerable intentionally or unintentionally occurs. These vulnerabilities exist in different forms in all levels and links of the information system, and can be utilized by the malicious actors, thus affecting the normal operation of the information system and its services.

Main Security Vulnerability Detection Technology

At present, the software security analysis methods are mainly divided into two categories: static analysis [2] and dynamic analysis [3]. Dynamic analysis is a test case as an input, and then run the executable file, run through the

Advances in Materials, Machinery, Electronics II
AIP Conf. Proc. 1955, 040054-1–040054-4; https://doi.org/10.1063/1.5033718
Published by AIP Publishing. 978-0-7354-1654-3/$30.00

output of the program to find defects in the process; static analysis without running the program under test cases directly reviews the analysis of the source code to find defects in the program.

At present, the commonly used method of detecting traditional security vulnerabilities is hybrid detection [4], that is, a combination of static analysis and dynamic analysis, as shown in the following figure1:

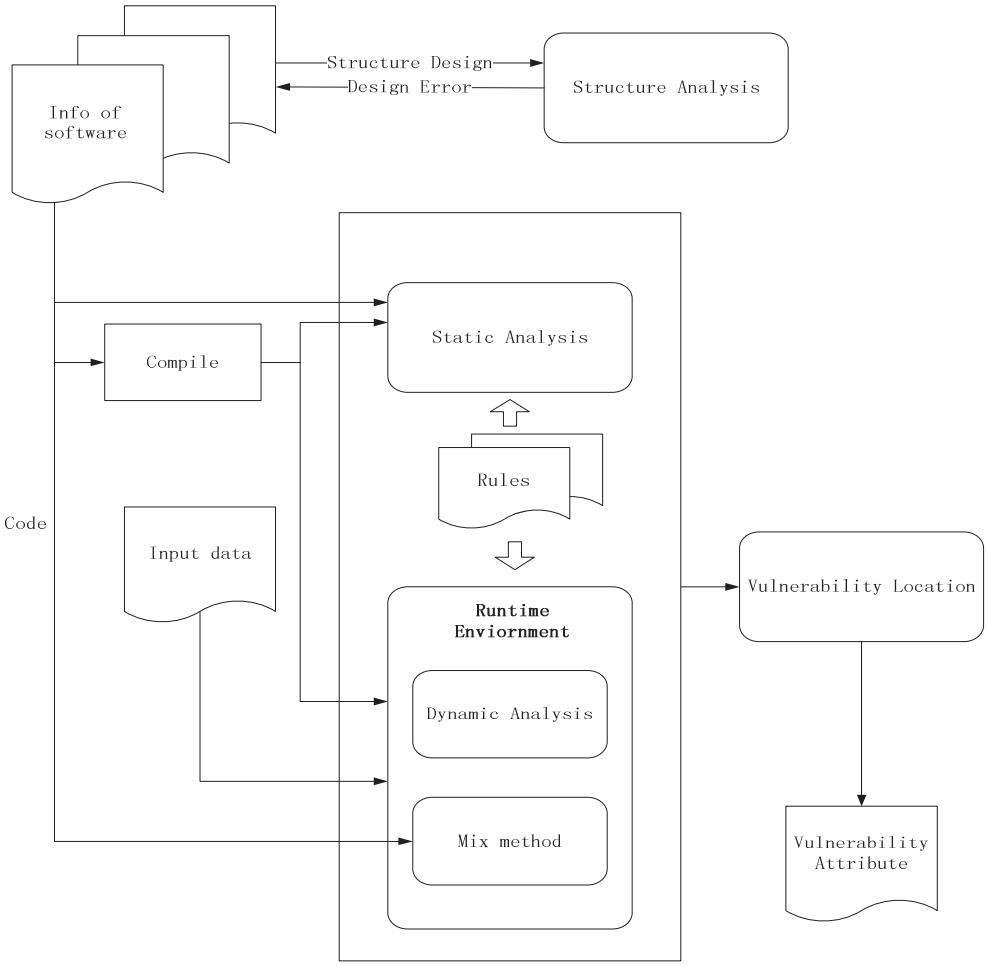

FIGURE 1. Normal Vulnerability Detection Framework

MACHINE LEARNING APPLICATION IN THE FIELD OF SECURITY

Machine learning methods have practical application in software defect finding, malicious code detection and intrusion detection. The use of machine learning method of the application framework as shown below figure2:

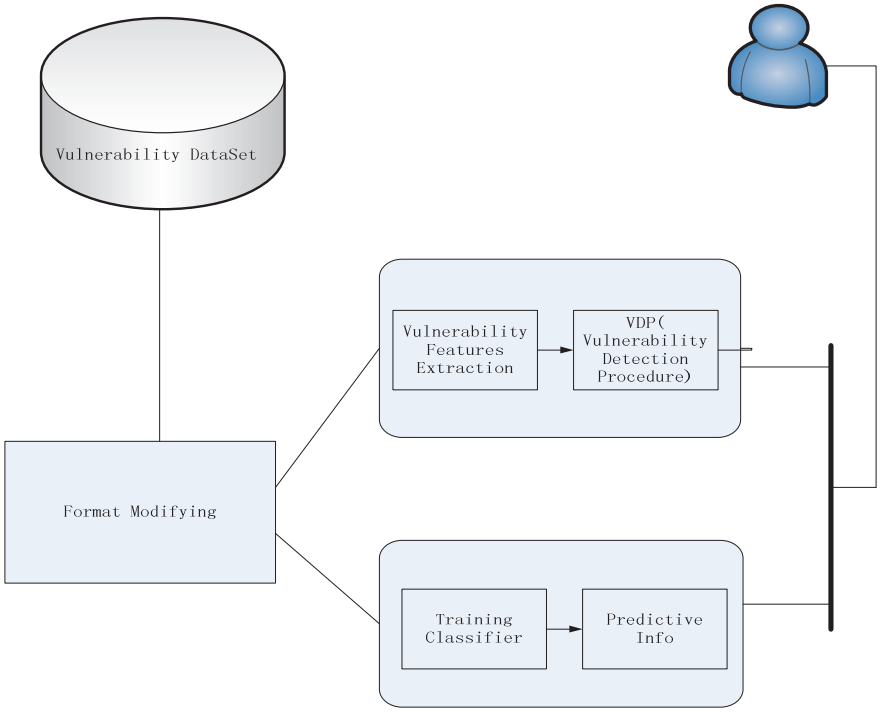

FIGURE 2. A Usual Detection Framework using Machine Learning

Intrusion Detection Based on Machine Learning Methods

The purpose of intrusion detection is to separate the intrusion behavior data from the normal data as far as possible. The traditional detection methods are mostly based on statistics and rule-based, and combined with machine learning Intrusion detection algorithms can be mainly divided into neural network-based intrusion detection methods and intrusion detection methods based on data mining.

Intrusion detection method based on neural network uses neural network framework to identify events of intrusion. As is known to all, neural network method with its parallel computing, distributed storage, and multi-layer structure, is suitable for computing large-scale, high-dimensional network data. The advantage is that it can handle large-scale, high-dimensional data, and the disadvantage is that the built-in neural network hidden layer topology and the output results are usually difficult to explain.

Intrusion detection method based on data mining uses data mining approach to find abnormal moods. Classification and clustering analysis in data mining are usually used for the identification of attacks. Techniques such as association rules analysis are suitable for research on complex network attacks. In recent years, many teams are dedicated to the study in this direction.

Machine Learning Used for Defect Discovery

Defect finding and vulnerability targeting in the software account for a large percentage of the project cost. However, the cost of defect handling can be greatly reduced by early introduction and improvement of defect prediction methods. In addition, this information can be used as feedback on program development to speed up the process of development. For these reasons, the use of machine learning methods to predict software defects can better serve the needs. Machine learning-based early defect prediction [5] has proven to be effective for obtaining high-reliability software.

Principal Component Analysis for Defect Finding

Using principal component analysis (PCA) to find the source code defects combines mining technologies, such as text mining. This approach embeds code into vector spaces and leverages machine learning to automate the detection of API usage patterns. From a known flaw, these patterns can be used to guide code reviews and identify potentially flawed codes with similar characteristics.

Support Vector Machine for Defect Finding

Using a public documented database of past vulnerabilities, a support vector machine (SVM) [6] is trained and used to predict if and how long a vulnerability may be exploited. In terms of input, the SVM works on high-dimensional vectors extracted from text fields, time stamps, cross-references, and other entries in existing vulnerability disclosure reports. Compared with the current industry-standard heuristics based on expert knowledge and static formulas, this SVM classifier is more accurate in predicting whether and when a vulnerability is likely to be exploited.

An Improved Security Vulnerability Detection Method Based on Machine Learning Algorithm

The machine learning algorithm and model matching theory are used to extract the characteristics of security vulnerabilities and to train effective classifiers. Then, according to the model matching knowledge, the vulnerabilities to be detected are judged and predicted. Model testing is a formalized method of checking whether a system satisfies a given property by exhaustive search of the state space. Model detection methods usually represent the nature of a program to indicate whether a state of the software system is reachable. Adopting the idea of model matching, through the collection of loopholes knowledge model of meta model specification, the design and integration of loopholes-oriented knowledge base are completed. According to the existing detection techniques based on pattern matching, this paper introduces the machine learning method in theory and puts forward a method to discover the security vulnerabilities based on the matching of the vulnerability models.

SUMMARY

Many new theories and new technologies emerge in the direction of machine learning in the field of artificial intelligence, exuding great magic. The traditional fields of security need to inject fresh blood. We need to combine the new theories and technologies with the requirements of the times, advance the research on the basis of the existing ones, and meet the security needs in the new era and the new situation. Based on the analysis of related research on security vulnerabilities and the summary of the current application of machine learning methods in the field of security, this paper argues that the research on security vulnerabilities detection based on machine learning method will be a field that deserves further study and a fertile land that has not been reclaimed land.

REFERENCES

1. Cui. An International Chapter of Internet Security Incidents in 2015 [J]. Confidentiality Science and Technology, 2015 (12): 26-27.
2. XIA YiMin, LUO Jun, ZHANG MinXuan, et al. Security Vulnerability Detection Study Based on Static Analysis [J] .Computer Science, 2006, 33 (10): 279-282.
3. Sun Jian. Dynamic detection of computer security vulnerabilities [J]. Information and Computers: Theory, 2013 (3): 25-25.
4. Zhang Junxian, Li Zhoujun. Research on Key Techniques of Security Vulnerability Detection Based on Mixed Symbol [C] // Information Security Vulnerability Analysis and Risk Assessment Conference. 2014.
5. MA Ying.Research on software defect prediction based on machine learning [D]. University of Electronic Science and Technology of China, 2012.Journal of Computer Applications 2015, 32 (4): 1145-1148.
6. Zhang Peng, Xie Xiaoyao. Research on Vulnerability Classification of SVM Based on Fuzzy Entropy Feature Selection Algorithm [J]. Application Research of Computers, 2015, 32 (4): 1145-1148.