Research on the Data Mining and Pattern Recognition Algorithms and the Applications on Information Security Engineering

Zhiyan Zhang,[1] Xinhua Zhang[2]

[1]Department of software, Anyang normal university, Anyang, Henan, 455000, China;

[2]Department of Computer, Wuhan Vocational College of Software and Engineering, Wuhan, Hubei, 430205, China.

Keywords: Data Mining, Pattern Recognition, Information Security, Engineering, Application.

Abstract. In this paper, we conduct research on the data mining and pattern recognition algorithms and the applications on information security engineering. The complete solution for network security firewall defense should have not only, should also be real-time monitoring to the network as can find intrusion behavior, and make the appropriate response. It is in this situation, the intrusion detection becomes the hotspot in the field of network security research. Under this background, we integrate the techniques of data mining and pattern recognition to enhance the traditional IDS systems with better implementation of the robustness and efficiency that is positive.

Introduction

With the rapid development of information technology and Internet technology is widely used, now network has become an indispensable part of people life, people's demand for information network system and dependence are increasing. At the same time, the threats to network security have become more and more serious. Therefore, this paper analyzes the reasons of influence of network security, and puts forward corresponding countermeasures of the network security becomes very important. In order to solve the problem of data security of the information network, the data system to conduct a comprehensive, reliable, safe and multi-level backup is essential, in addition to this, all kinds of basic security products, regardless of the firewall, antivirus, prevent hackers, prevent invasion and so on, are also able to protect data which could be organized as listed aspects.

- Unauthorized access. Refers to in advance without authorization, the system of corresponding access to the network or a computer's resources. Namely: try to avoid the basic system access control permissions, for the illegal use of network resources.
- Information disclosure. Refers to the valuable or sensitive data in intentionally or accidentally leaked out or lost, it includes information disclosure or on the transmission loss or leak in the storage medium [1-2].
- Destroy the data integrity. Refers to by illegal means get to the right to use the data, delete, modify, insert some important information, in order to obtain benefits the attacker's response.
- Malicious code. This kind of attack is likely to make the system perform specific procedures, cause serious damage, mainly including viruses, worms, spyware, trojans and other back door.

Information security risk management has become a mainstream paradigm of information security work, it embarks from the safety assessment, locate in safety control, the risk assessment theory and method used in information system, in the information system of the whole life cycle of a cyclical, and to evaluate the safety of information assets, scientific analysis of information and information systems in the confidentiality, integrity, availability, and so on security risks faced by security properties.

Under this condition, in this paper, we conduct research on the data mining and pattern recognition algorithms and the applications on information security engineering. Intrusion detection technology is the several key points in a network or computer system to collect information and analysis, find out

whether there is a violation of security policy and the signs of being attacked, and timely report system with unauthorized access or anomalies. In the later parts, we will discuss in detail [3].

The Proposed Methodology

The Principles of the Pattern Recognition. Pattern recognition is a kind of with the basic help of the computer, information processing, the discriminant classification process. Sentence classification in the application of scientific research and production practice is quite widespread, but often because of the impact factor of basic processing required too much, too complicated, to research and solve the difficulties. Pattern recognition makes people under the condition of the many factors that influence can still be convenient to many information processing, that using computer technology to summarize data, looking for a link between the target with a number of the factors, optimizing direction or target optimization area, to solve the practical problems has guiding significance and application value, and widely used, and achieved the greater success.

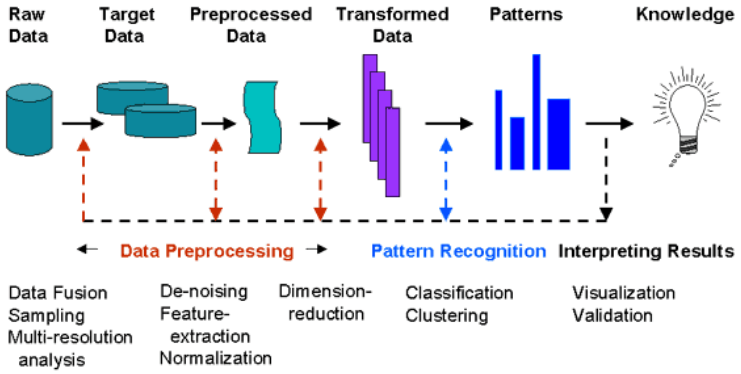

Fig. 1 The General Procedures of the Pattern Recognition

According to different objects and the different purposes, we can use different pattern recognition theory, method, the current mainstream technology and methods can be summarized as the following categories. (1) Statistical pattern recognition. This kind of recognition theory more perfect as method are many, usually more effective which has now formed a complete system. Although many methods, fundamentally is to take advantage of all kinds of distribution characteristics. (2) The syntactic pattern recognition. In many cases, for the more complex objects with some numerical characteristics can not only is more fully described, syntactic recognition technology can be used at this time. (3) The neural network. In terms of pattern recognition, and the method is significantly different after training the neural network is one of the specific treatment of general pattern feature extraction and classification recognition in the network can be done together. (4) The method of logical reasoning. It is a kind of the tied with statistical pattern recognition, syntax pattern recognition, intelligent pattern recognition method based on logical reasoning. It mainly includes the knowledge representation and knowledge acquisition that the knowledge reasoning three links [4].

Intrusion Detection. Intrusion detection is an important part of information security technology, is a kind of post-processing scheme, with intelligent monitoring, real-time detection, characteristics of dynamic response that relatively easy to configure and other measures to prevent the invading, IDS attaches great importance to by people gradually. As a new study, many computer experts have some simple computer immune model is developed. These results prompted expert further research on the immune system. Many researchers by feature selection to solve this problem, the number of feature extracting and processing is one of the most important factors to lead to a decline in too much. There is no linear relationship between features and classifier performance, when more than a certain limit, causes a classifier performance variation as the formula one.

$$\max_{s} G(S) \qquad s.t.s \in Separation \tag{1}$$

Feature selection can be seen as an optimization problem, and the key is to establish an evaluation standard to distinguish between features combination is helpful to basic classification, which features

combination redundancy, partially or completely unrelated as different evaluation functions may give different results based on the evaluation function to the classifier.

An IDS can be simply described as: dynamically monitoring behavior in a particular environment, and determine the behavior is the legitimate use system or an attack. An IDS generally use the three kinds of the information: the ID technology related information for a long time; System of the current configuration information by the protection system of audit information. According to information, the possibility of IDS is to evaluate the system behavior.

Let abuse recovery module work with scheduling module to process analysis module performance degradation or failure, choose other analyzer continue to analysis the problem of task that to ensure the robustness of the whole system when the data collection and pretreatment module to the network data flow as submit the request to the scheduling module. Scheduling module, first of all, according to the resource requirements description information in the core resource information database search to provide corresponding service resources, and then according to certain resource allocation algorithm provide service for the data analysis module, and return to the related information to data collection and pretreatment module. This set of rules is security experts on the analysis of the intrusion behavior experience and rules to the invasion of the past, the system is based on the known weaknesses, with a security policy. By the protection system of audit events are translated into facts of expert system and reasoning machine using these rules and the fact that the conclusion.

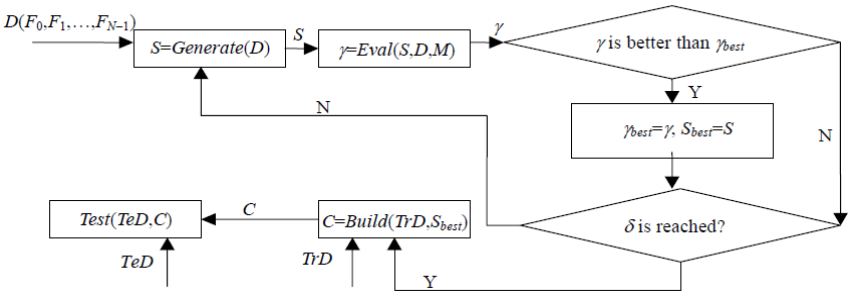

Fig. 2 The Inner Architecture of the Intrusion Detection System

Data Mining Techniques. Big data is different from the traditional "big data", the corresponding improvement of data mining technology is also necessary and the traditional single or simple cluster distributed mining algorithms that already cannot adapt to characteristics of big data while distributed mining algorithm can to a certain extent to make up for the deficiency of the existing algorithm.

Data mining is a new cross subject and also is inevitable result of modern science and technology mutual penetration, its basic goal is extracted from a large amount of data hidden, potential and useful knowledge and information. Due to data mining can build competitive advantage for the enterprise, bring huge economic benefits for the society, some well-known international companies have joined the ranks of the data mining, research and development related software and tools. Accordingly, we should summarize the procedures of mining as follows [5].

- Data selection. This is for the goal of knowledge discovery search and selects the relevant data including the unity of the different mode of data transformation and data and summary.
- Data preprocessing and data conversion. Their main task is to clean the wrong or conflicting data as will be more data in a file or database runtime environment merge processing, to avoid data inconsistency or redundancy.
- Data mining. The KDD is one of the most important stages as it'll be the first to determine the mining task or purpose, and then decided to use what kind of mining algorithm, and the actual mining operation, find useful patterns or knowledge from the database.

Data mining system framework is generally made up of three basic parts: data preparation system, modeling and mining system, the interpretation and evaluation system. In practice, however, there is no obvious boundary of the three systems because the process of data mining is a process of repeated cycle, from the data pretreatment, data mining, model was established, to the evaluation results as can be from any of the steps to return to the previous link.

Information Security Engineering. Although computer network information security threats, to take appropriate protective measures we can effectively protect general network information security commonly used computer network information security protection strategy.

Between network firewall technology is a kind of used to strengthen the network access control, to prevent to illegal means to enter the internal network, external network users access to the internal network resources, protect the internal network special network interconnection equipment operating environment. Intrusion detection as a new type active network security protection technology, or as protected by monitoring network system, the correctness of the test system configuration and security holes, statistics and analysis of the abnormal behavior model.

The basic idea of basic data encryption is to change the arrangement of information or replace or replacement according to certain rules, only legitimate and can understand the content of information, confidentiality of information. Address translation type firewall to inside of the IP address translation into temporary, external, the IP address of the register, internal network access to the Internet, to hide the real IP address. When a client needs to use the data on the server, first of all to send data requests to the proxy server, proxy server and then according to the request to the server for data, and then transfer to the client. Due to the external system with no direct data channel between internal server and external malicious abuse it is not easy to damage to the internal network system.

Fig. 3 The Components of the Information Security Engineering Subject

Summary and Conclusion

In this paper, we conduct research on the data mining and the pattern recognition algorithms and the applications on information security engineering. The existence and development of the IDS and the rapid development of the Internet is inseparable, Internet related services to get a promotion, IDS is important link and he current IDS development still faces many challenges, such as: attack feature extraction has not unified standard and the extraction of characteristic pattern library and update also depends on the manual way. In the future research, we will enhance the information security based countermeasures with more new methods and approaches.

References

[1] Siponen, Mikko, M. Adam Mahmood, and Seppo Pahnila. "Employees' adherence to information security policies: An exploratory field study." Information & management 51.2 (2014): 217-224.

[2] Cezar, Asunur, Huseyin Cavusoglu, and Srinivasan Raghunathan. "Outsourcing information security: Contracting issues and security implications." Management Science 60.3 (2013): 638-657.

[3] Laszka, Aron, Mark Felegyhazi, and Levente Buttyan. "A survey of interdependent information security games." ACM Computing Surveys (CSUR) 47.2 (2015): 23.

[4] Tsohou, Aggeliki, et al. "Analyzing trajectories of information security awareness." Information Technology & People 25.3 (2012): 327-352.

[5] Järveläinen, Jonna. "Information security and business continuity management in the interorganizational IT relationships." Information Management & Computer Security 20.5 (2012).