Approach Research on the Techniques for Network Intrusion Detection Based on Data Mining

Gong Lina[1,a] ,Xu Tao,Zhang Wei,Li XuHong,Wang Xia,Pan Wenwen

[1] The University Of Zaozhuang, Zaozhuang 277160, ShanDong,China

[a]390701387@qq.com

Keywords: Network Anomaly; Network Intrusion Detection; Data Mining;

Abstract: Along with computer technology's popularization and application and popularization, the network technology has been widely used, the resulting network security issues have become increasingly prominent, the network itself and network information system which is based on the potential security risks. Expounds the concepts of intrusion detection and data mining, common intrusion detection techniques and models, and analyzes the data mining technology in intrusion detection system application and optimization provide reference and perfect network intrusion detection system and reference.

Introduction

With the computer network and globalization, the social each domain in the Internet era has a qualitative leap, people's learning, work and life are integrated into the network, people through the network to share resources. After decades of development, in the network environment in the great changes, from the simple to the complex structure, a variety of network technology extension in time and space, equipment and the increase of the number of users, frequent network attacks, such as the network unstable factors difficult to make network management [1]. Only computer network security, information society's normal development, the national information security, to ensure network life is not violated, the people therefore, research on network security technology has important social significance and practical significance.

Intrusion detection is a hot technology in the network security research and achieved a certain development, also launched a number of commercial products, but because of the detection efficiency is not high, do not have a strong adaptability, and disadvantages, such as lack of can be extended is difficult to fully meet the needs of computer network security, also need to continue to learn and make the intrusion detection system more perfect [2]. Therefore, data mining technology research, in this paper, the design and realized a new data mining model and applications in network intrusion detection system, improve the existing defect detection algorithm and model, to improve the efficiency of the intrusion detection system.

The basic technology of data mining

Data mining is a long-term research and application of database technology is an inevitable result of the development of database technology but also to a more advanced stage, it cannot only large amounts of historical data query, data can also be found in the history of the unknown potential link [3] . Faced with the current state of large amounts of data, the association rule is an important branch of data mining, as the study an advanced and intelligent data processing and analysis technology has become a hot spot. By association rule mining, the amount of useful information can be implied in the sea there is the potential value of the data [4]. Target association rules is an effective method to extract the most interesting patterns. So far, it has been proposed many effective association rule mining algorithm, suggesting the algorithm is proposed mining algorithm most famous Agawal algorithm is based on a priori, but it is efficient in terms of time and space scales are faced with the challenge Therefore, many researchers explore new mining method, expanding the concept of association rules and applications.

Data mining refers to the large amount of data warehouse, reveals implicit, previously unknown, potentially important process valuable information, data mining is a decision support process, which is mainly based on artificial intelligence, machine learning, pattern recognition, statistical, database, visualization techniques, highly automated analysis of enterprise data, make inductive reasoning, dig out potential model to help the user to adjust the marketing strategy, reduce risk, make the right decisions. From the perspective of the process of data mining techniques, as shown in figure 1.

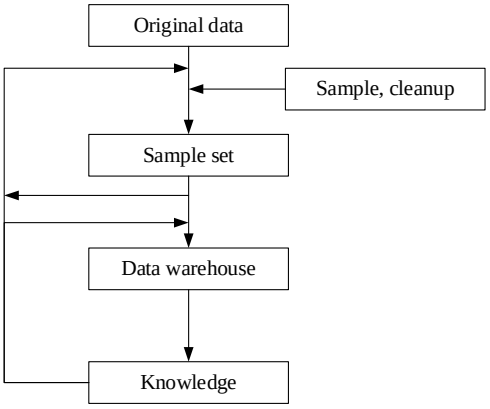

Figure 1.The basic process of data mining

Bayesian network based Theory for network intrusion detection

After data collection, sampling and cleanup needs. Cleanup is the result of the sample data set. Data Warehouse is an effective form of data storage, data mining is very beneficial. You can use a variety of data mining algorithms. Sometimes, the need to return to the final stage of the above process.

This is a directed acyclic graph, each node represents a random variable and an edge said direct probabilistic dependencies between two connected nodes. For each node, contains the node has a conditional probability distribution probability of the different values in the value of his parents. Formal, lattice structure assertions, each node is conditionally independent of all non-descendants to its parent node. The probability is shown in the formula 1.

$$p\left(X_1, X_2, ..., \mathrm{X}_n\right) = \prod_{i=1}^{n} p\left(X_i \mid Pa_G\left(X_i\right)\right) \tag{1}$$

Model resolution is the choice of training data sparseness of lack of scoring index, namely the size of the data set, small relative to the number of variables. In this case, there can be many different BN the same training data fitting. Therefore, using a single BN can lead to bad data to predict the future.

A promising solution to alleviate this problem is to employ BMA, which provides a principled approach to the model uncertainty problem by integrating all possible models weighted by their respective posterior probabilities. The prior probability is shown in the formula 2.

$$p\left(x \mid D\right) \approx \frac{\sum_{G \in g} p\left(x \mid G, D\right) p\left(G \mid D\right)}{\sum_{G \in g} p\left(G \mid D\right)} \tag{2}$$

The detection rate and false alarm rate we use to evaluate the performance of the algorithm to detect network intrusion. It is necessary to pay more attention to the false alarm rate, because in a real application, most of the network behavior is normal. High false alarm rate waste of resources, because each alarm must be checked. Adaboost-based learning algorithm of detection rate and false alarm rate depends on the initial weights of the training sample. So we propose to adjust the initial sample weights in order to balance the detection rate and the false alarm rate. The formula 3 calculate the solution.

$$R_c = C_c + \sum_i p_i \left(d_c^i A_c^i + \left(1 - d_c^i\right) A_m^i\right) \tag{3}$$

The design of intrusion detection system based on data mining

In the algorithm of data mining algorithm for mining association analysis can be found that the network connection data attribute, the relationship between sequence analyses algorithms can be found about invasion associated characteristics of intrusion attack. Through sequence pattern analysis method is applied to obtain the intruder behavior sequence relation, be able to get intrusion behavior characteristics of the temporal information, similarly also can get the characteristics of the normal behavior, according to the time sequence characteristics of user behavior to determine the user's behavior is normal behavior or intrusion behavior. Using correlation analysis algorithm and sequence analysis algorithm to construct the normal patterns of behavior and applied to anomaly intrusion detection; finally, the algorithm carries on the classified analysis, can from the training data obtained from mining to identify normal behavior and intrusion behavior rules. The principle and the mining process is shown in figure 2.

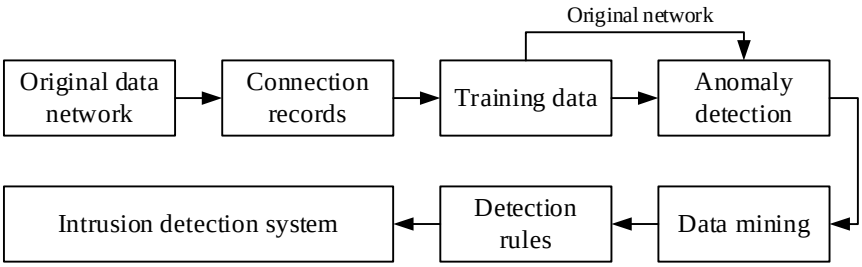

Figure 2. The mining process of intrusion detection system

Intrusion detection system model based on data mining framework basically has the following several parts: data preprocessing module, association rules mining module, misuse detection rules mining module, as shown in figure 3.

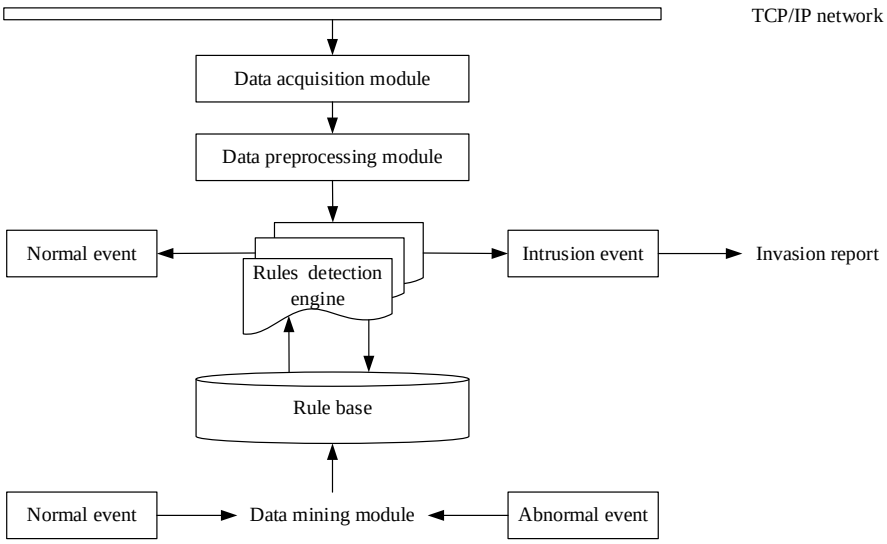

Figure 3. The intrusion detection system model based on data mining

Data acquisition module. Mainly responsible for efficiently intercepted packets on the network, the efficiency is the main factor. The module data packets intercepted by the submitted to pretreatment module, as the input of data sources. The module is the "eyes" of the whole system, it is about whether network data access to timely, comprehensive and reliable, will directly affect the efficiency and accuracy of the whole system.

Data preprocessing module. The module in the data acquisition module for data source, and through to the data source data format, encoding operations, such as the irregular data source information into standard data information system can normal use. The module of intercepted data explanation effect, through the module system of transformation to generate data that can be recognized.

Rule detection engine. the detection engine main execution of intrusion detection based on rules, will be formatted data source information to match the rules of rule base, according to the matching results to determine whether there is intrusion events.

Rule base. storage system all deposits and artificial join the invasion of the rules. At the same time in order to avoid the rule base too much influence the efficiency of detection, rule base is also responsible for their own maintenance on a regular basis.

Normal event store. store all rules detection engine to a normal event features. These special field will be in the form of item sets as input of the data mining module, so that the whole system adopts data mining algorithm to get the connection between each set, and in the form of rules into the rule base of the whole system.

Abnormal events library. To store all the rules for abnormal events detection engine characteristics. The event storage way and its role in normal event store.

Data mining module. This module with the improved Apriori algorithm are not present in the mined rules library rules, filtering to mislead users to update the rule base. Due to the module USES is association rules in data mining method, the "knowledge" in the form of rules mined performance to decision makers.

Conclusion

With the rapid development of computer network, various problems also arise, especially the network security problem. How to quickly and efficiently find a variety of network intrusion behavior, to ensure the safety of the system and network resource is very important. Intrusion detection is a proactive safety protection technology, to take the initiative to monitor and track the invasion, to protect the computer system, network system, and the safety of the information infrastructure has very important practical significance. In this paper, in light of the characteristics of the network environment, this paper proposes a new network anomaly detection based on data mining technology, the method of clustering analysis in data mining to extension, raise the efficiency of intrusion detection system, guaranteeing the security of the network..

Reference:

[1] P. Garcia-Teodoro, J. Diaz-Verdejo, and G. Maciá-Fernández: Computers & security, Vol. 28(2014) No.1, p. 18.

[2] M.A. Aydın, A.H. Zaim, and K.G. Ceylan: Computers & Electrical Engineering, Vol. 35(2012) No.3, p. 517.

[3] W. Hu, and S. Maybank: Systems, Man, and Cybernetics, Part B: Cybernetics, Vol. 38(2010) No.2, p. 577.

[4] A. Boukerche, R.B. Machado, and K.R.L. Jucá: Computer Communications, Vol. 30(20011) No.13, p. 2649.

[5] J. Kim, P.J. Bentley, U. Aickelin: Natural computing, Vol. 6(2008) No.4, p. 413.