博士論文概要

論文題目

Study on Intrusion Detection using Average Matching Degree Space based on Class Association Rule Mining

申請者
Nannan LU

情報生産システム工学専攻
ニューロコンピューティング研究

2012年11月
Intrusion detection system is designed to detect actions that attempt to compromise the integrity, confidentiality or availability of resources. With the increasing number of users and systems connected to the networks, intrusion detection system is an essential component in the overall network and information security.

Traditional intrusion detection techniques have misuse detection and anomaly detection. Misuse detection utilizes the known patterns of previously seen attacks to detect intrusions. However, it easily fails to recognize brand new attacks. While, anomaly detection aims at identifying intrusions by looking for the deviation with known normal patterns. In actual systems, however, anomaly detection results in a large number of false alarms. Anomaly detection is also difficult to be realized in highly dynamic environments. Various methods aim at improving the detection ability of intrusion detection systems. Among these methods, data mining is still a safe and efficient method for intrusion detections. In this thesis, data mining generally refers to the process of discovering useful rules from a set of historical data.

Class association rule mining is used in this thesis to present the normal behaviors or misuse intrusion behaviors. Therefore, the framework of the intrusion detection system proposed by this thesis contains three essential components which are rule mining, rule pruning and classification. This thesis proposes a series of new methods in terms of the three topics, in order to detect a wide variety of attacks and avoid a large number of false alarms.

As the basis of the whole system, an evolutionary method is used to extract the class association rules for the proposed intrusion detection systems. As an extend evolutionary algorithm of GA and GP, Genetic Network Programming (GNP) has been successfully applied to a wide range of real life applications, including elevator supervisory control, buying and selling strategy in stock market, traffic prediction and intrusion detection. GNP mainly extracts useful rules from historical network connection data for intrusion detection systems. Then, the average matching degree is proposed to evaluate how much the data matches with the rules in normal class and misuse intrusion class. By this way, the multi dimensional problem is projected into a two dimensional average matching degree space. Next, this thesis focuses on improving the quality of the rule pool by pruning
redundant and irrelevant class association rules. The final task is how to build efficient classifiers.

Chapter 1 briefly introduces the research background of intrusion detection and discusses related researches on this field. In addition, the motivations and contributions of this thesis are described. Furthermore, the thesis organization is presented in this chapter.

In chapter 2, a GNP-based data mining framework for intrusion detection is proposed to utilize the advantages of both misuse detection and anomaly detection. In this system, GNP undertakes extracting class association rules, especially it can extract many interesting and important class association rules from network connection data efficiently using the sub-attributes based on information gain. Information gain can avoid the information loss as much as possible when dealing with the partition of continuous attributes. These rules are evaluated by the average matching degree of the data with rules. In this way, the multi-dimensional data space is converted into a two-dimensional space. Then, a classifier is also proposed by combining misuse detection and anomaly detection to classify the new data. By combing misuse detection and anomaly detection, the simulation results show that the proposed system has better performance.

However, two problems have come out. One of them is that too many rules bring much useless information into the rule pool and waste much processing time during classification. In this case, a pruning method is needed to reduce the useless rules from the rule pool. The other problem is that although information gain can avoid the loss of much information, but the discretization of the continuous attributes into intervals will lead to the sharp boundary problem. Therefore, in chapter 3 and 4, we focus on solving the two problems, respectively.

Chapter 3 proposes an efficient class association rule pruning method. This rule pruning method has two stages. The first stage can pre-prune the rules to improve the efficiency of the second step. The second step implements GA to pick up a small set of effective rules among remaining rules in the first stage. The simulation results show that the proposed two-stage rule pruning method can improve the detection ability of intrusion detection system under a small set of class association rules.
Chapter 4 proposes Fuzzy GNP to extract class association rules. Comparing with conventional GNP, Fuzzy GNP can deal with both discrete and continuous attributes in intrusion detections and overtake the sharp boundary problem of the sub-attributes method. Each continuous attribute has its own initial fuzzy membership function and its parameters are evolved with GNP evolution. In addition, the probabilistic node transitions replace the traditional node transitions in GNP, which can contribute to extracting diversified rules. The simulation results on KDD Cup 1999 show the good detection ability, where Fuzzy GNP can extract more useful rules than conventional GNP.

Chapter 5 and chapter 6 analyze the classifier for effective intrusion detection systems.

Chapter 5 proposes a Distance-based classification method. Firstly, K-closest neighbor classifier is employed to categorize each new data into either normal or misuse intrusion. Then, the centroids of anomaly intrusion data are defined by the centroids of normal data and misuse intrusion data. Distance-based classification method can distinguish anomaly intrusions from the mixture of normal and misuse intrusions more effectively than other methods. The simulation results on NSL-KDD data set indicate that the detection ability of anomaly intrusions has been improved. From the results on KDD Cup 1999 data set, it is remarkable that the Distance-based classifier can detect all the connection data of neptune type and smurf type with known patterns, while the detection performance of this classifier is not so sensitive to parameter K, which is the number of the closest neighbors from the new data.

Chapter 6 proposes a new classifier using Gaussian functions and clustering method. To make full use of normal and misuse intrusion patterns, the proposed method groups the similar patterns into the same cluster. Then, Gaussian functions are used to look for the boundary for each cluster, where GA is used to decide the shapes of Gaussian functions. This classifier can classify the new connection data as normal, misuse intrusion or anomaly intrusion fairly correctly. Especially it can distinguish normal and anomaly intrusions well.

Finally, chapter 7 summarizes the proposed methods in this thesis and presents the corresponding conclusions.