Application of hospital information security intrusion detection system based on the new NPRIM algorithm

Xingshan LI [a], Min XU

Luohe medical college, Luohe, 462000, China

[a]email:604141388@qq.com

Keywords: clustering, boundary point detection, Intrusion detection

Abstract. The specific model of intrusion detection based on clustering and boundary points detection was elaborateed in this paper, and the data processing, clustering analysis, intrusion judgment, intrusion response, typical data warehousing the five stages are described in detail; Then the experimental environment and the experimental results were compared and analyzed, and further validation based on the improved NPRIM algorithm of the project team applied to intrusion detection is effective and feasible.

Introduction

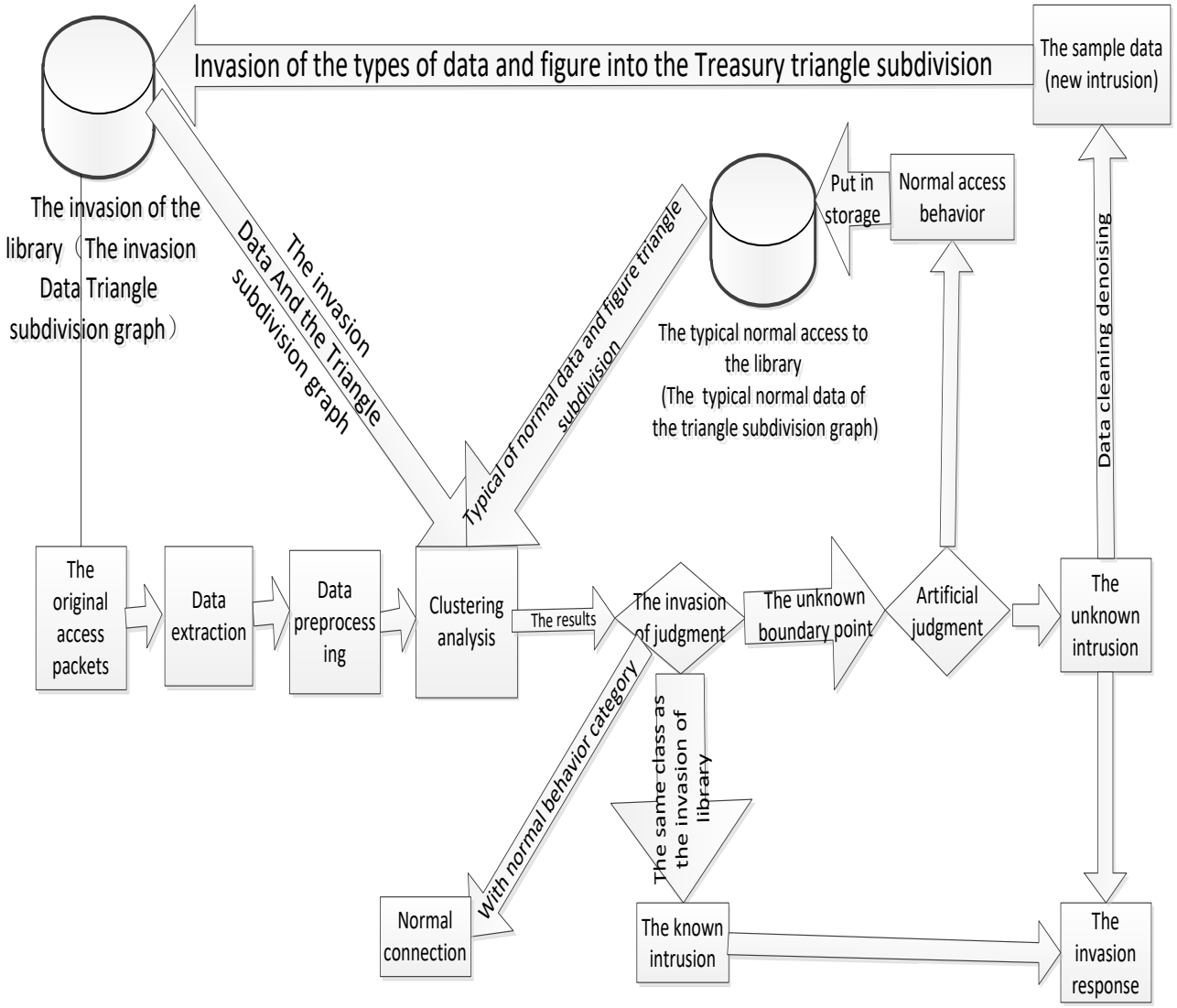

Fig.1 Integral flow chart of intrusion detection model

Design of Intrusion detection model based on clustering and boundary detection is, clustering and boundary detection NPRI algorithm were applied to intrusion detection system，The whole

process is shown in figure 1 .

Can be seen from the integral flow chart of intrusion model, the intrusion detection system of the project group is mainly composed of data processing, clustering analysis, intrusion detection, intrusion response, and storage five parts.

Data processing

Data processing includes the original access data packet, data extraction, data preprocessing three parts, Among them, access to data extraction is mainly through the extraction of data from the network to monitor and detect. For example, through the network layer of the IP package, the link layer data frame, etc. Data preprocessing is mainly to collect the network data pretreatment to make it standardized, it is transformed into a suitable data format, and the data to feature selection, filtering noise, filtering and other operations [1].The project team uses the classic KDD Cup 1999 data set to do validation experiments.

KDD Cup 1999 data set is the most widely used in the academic community to detect the data source of the intrusion detection system. This data set is the data collected by the DARPA authorized MIT Lincoln Laboratory in 1998 to simulate the real network attack environment.The data set is used for data mining, which has been recognized as the standard data set in the field of intrusion detection.

Cluster analysis

This part mainly is to use the improved NPRI algorithm through the establishment of the triangle subdivision graph was carried out on the processed data clustering. Depth first traversal from each of the non classified interior point P, the candidate boundary points adjacent to the internal points and the internal points are classified as a class, and each class is marked, and the results are saved. It is worth noting that, in order to reduce the artificial intervention, the abnormal data in the database, the typical normal data with the preprocessed data after clustering[2]. Add a Z parameter for each cluster, the parameters of the initial value is 0, in the process of clustering, and tagged intrusion data directly to a class of data marked as 2 said invasion behavior, 1 of the data that has been marked as a class of normal data indicates the normal data.

Intrusion judgment

This part mainly carries on the judgment to the result of the cluster analysis, which is composed of the automatic intrusion judgment and the artificial judgment. Automatic intrusion detection stage, the results of the cluster analysis and the database in the database of a class of data objects, that is, the variable Z value of 2 of the data, as a known intrusion to submit to the intrusion response, The results of clustering analysis and the typical normal access to data objects in a class of data objects, that is, the variable Z value of 1 of the data, as a normal behavior does not deal with. For the unlabeled "normal" and "invasion" of the data object, that is, the variable Z value of 0 of the data submitted to the administrator for artificial judgement[3].

Artificial judgment stage, In order to save time and energy, the system first submitted the data object ,which unmark "normal" and "invasion" located in the class Ci boundary points to the administrator to judge, and then the results were judged according to artificial judgment. If the boundary points of Ci are intrusion, all data objects in Ci are intrusion behavior, if the boundary points of Ci are normal access behavior, all data objects in Ci are normal access, If the boundary points of the Ci are both intrusion and normal access, the neighbor of the boundary point of the intrusion behavior will be submitted to the administrator, so that it can determine whether the data points in the neighbor set is the intrusion behavior. If it is, then continue to judge the neighbor data points, until the exception cannot find neighbors so far; if not, then marked as normal behavior. The intrusion behavior, which is judged by the artificial judgment stage, is unknown [4].

Intrusion response

Intrusion response is to effectively prevent the further occurrence of intrusions after the intrusion behavior and suspicious behavior were detected, and a series of response measures was taken for reduce the loss caused by the invasion to the minimum. Usually, intrusion response includes warning, recording, tracking, blocking, forensics, counter attack, loss recovery and so on.

The intrusion response is divided into three types, the notification type, the semi-automatic type and the automatic type. The project team uses the notification type of intrusion response, When an intrusion and a suspicious behavior was detected, the system sends an alarm to the security manager, the administrator according to the experience to take appropriate measures and measures. It is worth noting that, When Intrusion behavior and suspicious behavior are known intrusions, the system will provide administrators with the past, such as the invasion of processing records for the administrator reference.

Typical data warehousing

In order to quickly, effectively and adaptively detect the intrusion, the intrusion detection system designed by the project group has established two samples of the intrusion and the typical normal access database. Every new intrusion behavior is detected by the system, and the system will save the behavior and the related response of the security manager to the intrusion warehousing [5]. Similarly, the system can not automatically determine the normal access behavior of a system, the system will be stored in a typical normal access to the library. In the process of clustering analysis, intrusion database and typical normal access library two samples, with the detected data with clustering, which can improve the intrusion detection accuracy, but also reduces the workload of the security administrator.

Experimental environment, results and analysis

The experimental environment used by the project team is Intel（R）Core(TM) i3-3217 CPU @ 1.8GHz 1.8GHz, 4GB memory,Windows7 system platform was chosed,Matlab6.5 and VC++ language programming environment.

Typically, two performance indicators of detection rate and false detection rate was adopted to evaluate the intrusion detection algorithm, the definition of these two indicators:

Detection rate (DR)=

Detect the invasion of record number/The invasion of record total data set （1）

False rate (FR)=

The normal record number of false positives for invasion/ Normal record number of the data set （2）

These two performance indicators fully reflect the performance and capacity of intrusion detection, they can make a more impartial evaluation of the efficiency of the intrusion detection system: Among them, the greater the DR intrusion detection system detection is more sensitive, and the FR the smaller the intrusion detection system, the higher the correct rate.

In KDD Cup 1999 training data set, there are nearly 400000 attack data, accounts for about 80%, according to the number of objects. But in reality, the record of all of the invasion of the proportion between 1% and 1.5%. In order to avoid the leak and mistakenly identified data set must be filtered, Before test, select the record of all of the invasion of proportion between 1% and 1.5% of the data. The project team will be improved NPRIM algorithm and K-means algorithm performance comparison test, performance evaluation in the test. Six data set were chose from kddcup. Data _10_percent package as the experimental data object set in the experiment, among them, the data set contains a DOS, U2R, PROBING and R2L all four types of intrusion data types, and invasion of proportion of the total number of data sets of 1% to 1.2%. In the project, 3 discrete attributes of the

data set and 13 continuous attributes are selected in the experiment and used as the key attributes to cluster. Using the selected data set respectively in NPRIM algorithm and K - means algorithm to improve the experiment. Two algorithms for 6 times test clustering results are shown in table 1 below:

Table1 test results of intrusion data set

cluster number	detection rate		false drop rate	
	K-means	Improve NPRIM	K-means	improve NPRIM
10	18.9%	50.7%	0.42%	0.24%
15	32.7%	55.6%	0.46%	0.32%
20	42.8%	70.1%	0.53%	0.39%
25	52.3%	78.2%	0.68%	0.57%
30	55.2%	78.8%	0.86%	0.61%
40	56.5%	80.4%	1.09%	0.73%

You can see from the chart, improve NPRIM algorithm is obviously superior to K - means algorithm no matter in the detection rate and false detection rate. With the increase of number of clustering, two algorithms of detection rate and false detection rate is increased. Because when the number of clustering is very few, exception class rarely, normal record is not easy to be assigned to the exception classes, on the contrary, abnormal records will be easily assigned to normal class, thus detection rate and the false drop rate are low. When the clustering entries are increased, the abnormal class also increases, which leads to more normal records are divided into the exception class, while the abnormal records are more likely to be marked out, so that the detection rate and false detection rate at the same time were improved.

48 sub data sets was selected from the test data set, and the single intrusion test set is tested. The results of the experiment are as follows table 2:

Table 2 test results of single intrusion dataset

classification	Detection rate		false drop rate	
	K-means	Improve NPRIM	K-means	Improve NPRIM
DOS	61.3%	85.1%	0.70%	0.52%
U2R	46.4%	62.7%	0.78%	0.56%
R2L	31.7%	54.1%	0.92%	0.62%
PROBING	58.4%	78.6%	0.74%	0.48%

From table 2, we can see that the detection system can get the best results when the cluster number is about 25. Improved NPRIM algorithm of intrusion detection system detection results are significantly better than the detection effect of K-means algorithm, and the detection rate of invasion attack type of PROBING and DOSs are higher, but the invasion of the U2R and R2L invasion type attack detection rate are low. On the one hand, due to the less intrusion record of U2R and R2L of the test data set, resulting in poor detection results, on the other hand, due to the U2R and R2L types of intrusion data intrusion is masquerading as legitimate user identity attacks, so characteristic of its performance characteristics and normal network data packets are very similar, which resulted in lower detection rate.

The improved NPRIM algorithm has an advantage in the detection of a single intrusion attack type in terms of false detection rate. It is worth noting that many invasion records of R2L mingled with normal clustering in the training, which caused the R2L type of attack detection error rate is generally higher, and more normal record detection stage were judged as abnormal.

In the actual network environment, the DOS and PROBING types of intrusion attacks are the majority. Improved NPRIM algorithm to the invasion of known and unknown attacks are good detection effect is obtained in the paper. For the rare R2L and U2R in the actual environment, the improved NPRIM algorithm also has a certain ability to detect. Through the experimental analysis,

it is found that the improved NPRIM algorithm based on the project is effective and feasible for intrusion detection.

Summary

First elaborated the specific model of intrusion clustering and boundary detection based on detection and data processing, cluster analysis, judgement, intrusion response, typical data of these five stages are described in detail; then the experimental environment and the experimental results were compared and analyzed, further validation of the project team NPRIM algorithm based on the improvement in intrusion detection is effective and feasible.

Acknowledgement

In this paper, the research was sponsored by Medical Science Research project of Henan Province (Project No. 201404065).

Reference

[1] Kim G,Lee S,Kim S. A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), 2014: 1690-1700.

[2] Khor K C,Ting C Y,Phon-Amnuaisuk S. A cascaded classifier approach for improving detection rates on rare attack categories in network intrusion detection. Applied Intelligence, 36(2), 2012: 320-329.

[3] Davis J J,Clark A J. Data preprocessing for anomaly based network intrusion detection: A review. Computers & Security, 30(6), 2011: 353-375.

[4] Lin S W,Ying K C,Lee C Y,et al. An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection. Applied Soft Computing, 12(10), 2012: 3285-3290.

[5]Louvieris P,Clewley N,Liu X, Effects-based feature identification for network intrusion detection. Neurocomputing, 121,2013; 265-273.