Advances in Engineering Research, volume 104
7th International Conference on Mechatronics, Control and Materials (ICMCM 2016)

Intelligent intrusion detection technology NPRIM improved algorithm based on non- parameter clustering

Xingshan LI [1,a], Na LI[2], Min XU[1]

[1]Luohe medical college, Luohe, 462000, China

[2]Luohe vocational technology college, Luohe, 462000, China

[a]email:604141388@qq.com

Keywords: NPRIM algorithm, clustering, non-parametric clustering, boundary point detection, intrusion detection

Abstract. The specific model of Intrusion Detection Based on clustering and boundary points detection was elaborated in this paper, and the data processing, clustering analysis, intrusion, intrusion response, typical data warehousing the five stages were described. Based on the NPRIM algorithm, the clustering function is added, and the non-parametric clustering algorithm and the boundary point detection algorithm are proposed, the algorithm can not only cluster but also detect the boundary points. At the same time, in order to meet the needs of intrusion detection, the output will be processed accordingly. Further validation of the project is based on the improved NPRIM algorithm applied to intrusion detection is effective and feasible.

Introduction

Design of intrusion detection model based on clustering and boundary points detection, clustering and boundary detection NPRI algorithm was applied to intrusion detection system, the whole process is shown in fig1.

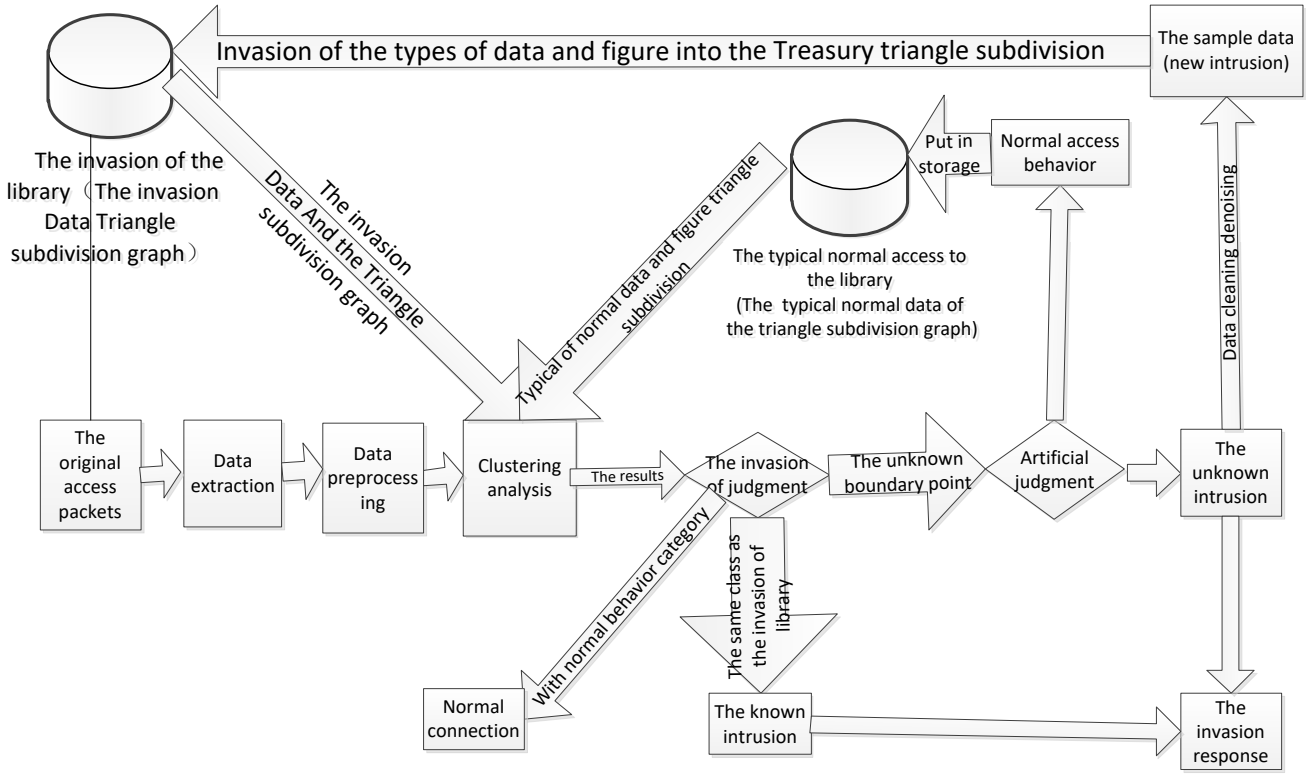

Fig1 Overall flow chart of intrusion detection model

Can be seen from the Fig1, the intrusion detection system of the project group is mainly composed of data processing, clustering analysis, intrusion detection, intrusion response, and storage of five parts [1].

Copyright © 2016, the Authors. Published by Atlantis Press.
This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

1). Data processing including the original access packets, data extraction, data preprocessing of three parts, among them, the access to the data extraction mainly through monitor and detection to extract the data from the Internet, for example, through the network layer IP packet, the data link layer frame, etc. Collected data preprocessing is mainly the network data preprocessing make its standardization, turning it into a suitable data formats, and the data to feature selection, remove noise and filtering , etc [2]. The project team uses the classic Cup KDD 1999 data set to do validation experiments.

2). Cluster analysis on the NPRI algorithm is mainly through the establishment of the triangle subdivision graph was carried out on the processed data clustering. The specific steps to follow the steps described above, in step 5, from every unclassified interior point P began depth first traversal, the candidate boundary points and interior points adjacent and adjacent interior points into a category, and for each type of marker, and save the results. It is important to note that in order to reduce human intervention, abnormal invasion repository data, typical normal after special tags with pretreatment of data clustering. In each cluster, a parameter Z is added, and the initial value of the parameter is 0. In the process of clustering, the data of a class of the intrusion data is labeled as 2, 1 of the data that has been marked as a class of normal data indicates the normal data [3].

3). Intrusion detection is mainly based on the results of cluster analysis, which consists of automatic intrusion detection and artificial judgment.

Automatically invasion stage, the clustering analysis results and invasion in the repository data object type of data, namely variable z value of 2 data, known as intrusion behavior is submitted to the intrusion response; The results of cluster analysis and typical normal access sample library data object in the type of data, is data variable z value is 1, that is, as a normal behavior will not be processing. 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 judgment.

4). Intrusion response is the intrusion behavior and suspicious behavior is detected, for the effective to prevent the further occurrence of the invasion, and the degree of loss caused by the loss to the minimum, a series of response measures taken. Usually, intrusion response includes warning, recording, tracking, blocking, forensics, counter attack, loss recovery and so on.

5). 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. Similarly, the system cannot automatically determine the normal access behavior of a system, the system will be stored in a typical normal access 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[4].

Improved NPRIM algorithm

On the basis of the NPRIM algorithm, the project group is added with the function of clustering, and the algorithm of non-parameter clustering and boundary points detection is proposed. At the same time, in order to meet the needs of intrusion detection, the output will be processed accordingly.

Improvement method

The data points in the data set usually contain three kinds: internal point, boundary point, noise point.

In combination with the related concepts of NPRIM algorithm, the project is defined as a clustering algorithm which is defined as the internal points and the boundary points which are connected with the interior points.

Can be seen from the boundary point detection algorithm of concrete steps, remove the noise

points at step 5, in the process of removing the first kind of noise points, from each of the unclassified internal point p begin depth-first traversal, the candidate boundary points and interior points adjacent and adjacent interior points into a category, clustering the marker number is less than $\sqrt{n}$ for noise. Thus, the NPRIM algorithm has the clustering but there is no clear mark on the data set, the project team will NPRIM algorithm step 5 from every unclassified interior point P began depth first traversal, the candidate boundary points of adjacent interior points and interior points adjacent and classified as a class, and for each class mark, will mark the clustering number is less than $\sqrt{n}$ for the intrusion, the results will be saved [5].

Improvement result

In order to verify the effectiveness of the project team to improve the NPRIM algorithm, the test was carried out on a number of comprehensive data sets, and 6 typical integrated data sets were selected, as shown in Figure 2.

Fig2 Typical integrated data set

In order to verify the effectiveness of the algorithm, the algorithm has been tested on a number of complex data set, we selected six typical integrated data sets, as shown in Figure 5:

（1）708 data object and five separate clusters in DS1, the distance between the two clusters is closer, and the density of a cluster is smaller than that of its adjacent clusters, which does not contain noise points.

（2）9993 data object in DS2, Which contains a number of noise points and four separate clusters of uniform density but different shapes.

（3）7832 data object in DS3, consists of two symmetrical diamond connected clusters and a small amount of noise points, two of them connected to the cluster density decreases from the middle to the side.

（4）5034 data object in DS4, which consisted of clusters with 5 arbitrary shapes, different sizes, different densities and a large number of noise, two of clusters connected with "short bridge".

（5）11680 data object in DS5, by a large number of noise and 6 arbitrary shape, different size and density of non-uniform connection cluster, and between each cluster are connected by a sinusoidal line noise.

（6）11399 data object in DS6, the data set contains 12 different shapes, different densities of clusters and a large number of noise point. The clusters are very close to the clusters and they are

connected by noise.

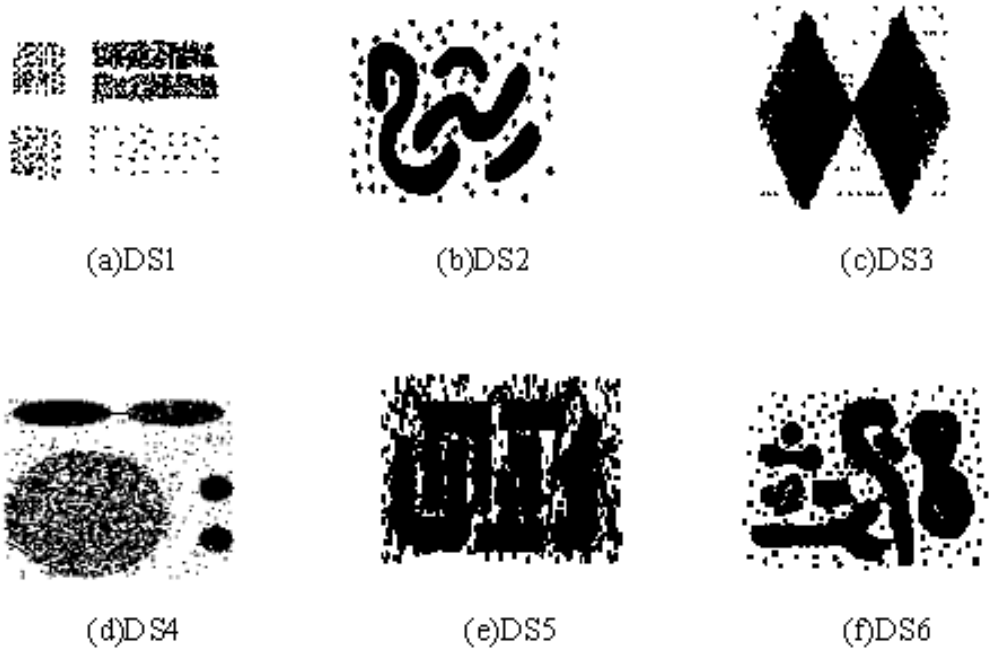

Fig5 Original data set

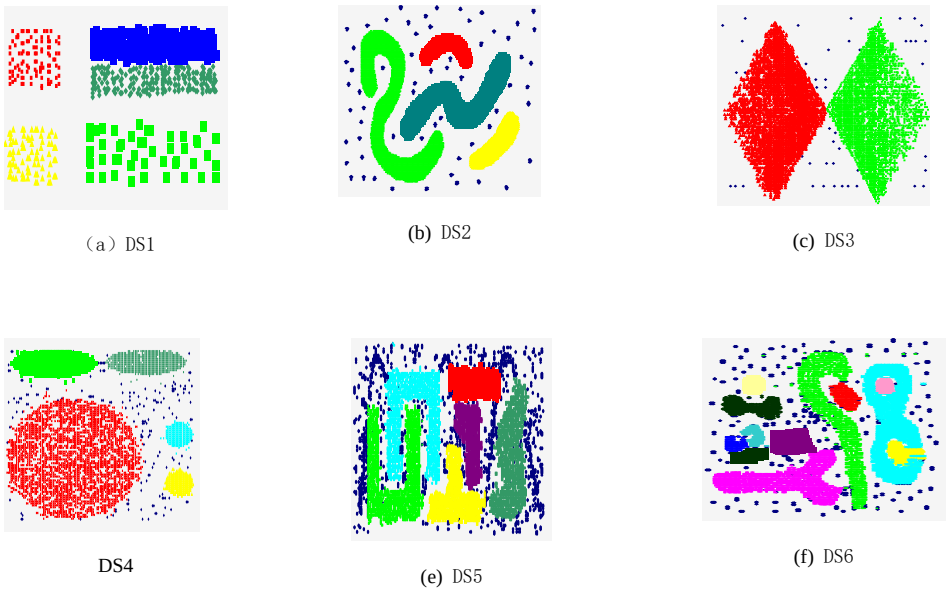

Fig 6 Clustering results of improved NPRIM algorithm

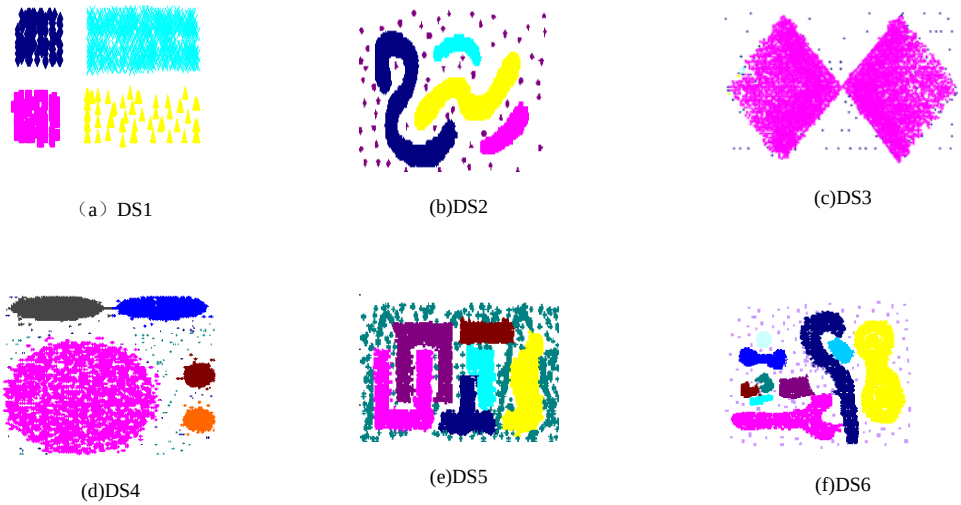

Fig7 Clustering results of SNN algorithm

Fig 6 and fig7 are the clustering results of the SNN algorithm and the improved NPRIM algorithm in the above six comprehensive data sets, the detection results of the two graphs are the best results of their detection on the corresponding data sets. Can be seen by comparing, as SNN algorithm by entering the global parameters to clustering, the change of density at intervals has no effect on the overall situation when the distance between clusters and cluster is close, therefore SNN suit to deal with uniform density distribution, large data sets, the distance between the clusters and cluster in non-uniform and contains links cluster data sets on its detection precision is not high. But NPRI algorithm with clustering functions in homogeneous and heterogeneous data sets can correctly identify the separation of arbitrary shape, size and density of clusters and connection clusters, and effectively remove the noise.

Summary

In this paper, first of all, the concept of the NPRIM algorithm and the algorithm of the boundary threshold calculation method, the specific steps of the boundary points detection algorithm and the detection results are described in detail. On this basis, according to the characteristics of intrusion detection technology, the NPRIM algorithm is improved, and the detection results are verified in detail, so as to prepare for the algorithm in the intrusion detection system.

Acknowledgement

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

Reference

[1] Ling Zhang, ZhongYing Bai, Shoushan Luo and so on. Based on raw sugar set and the integration of artificial immune intrusion detection model. Journal of communication. Journal on Communications. 34(9), 2013: 166-176.

[2]Li Y,Li J L,Yue S J,et al. Research of Intrusion Detection Based on.Ensemble Learning Model. Applied Mechanics and Materials,336,2013:2376-2380.

[3]Hwang T S, Lee T J,Lee Y J. A three-tier IDS via data mining approach. In;Proceedings of the 3rd annual ACM workshop on Mining network data. ACM,2007: 1-6.

[4]Fatma H,Mohamed L. A two-stage technique to improve intrusion detection systems based on

data mining algorithms. In: 5th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO). IEEE, 2013: 1-6.

[5] 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.