Volume 4, Issue 11, November 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Novel Intrusion Detection System Using Hybrid Approach

Punam Mulak*, Nitin R. Talhar
Computer Department, AISSMS COE,
Pune University, India

Abstract*— Intrusion detection system is used for securing computer networks. Lots of research is done in IDS. In this paper we discuss some data mining techniques for intrusion detection system. Different techniques have different accuracy and false alarm rate. Here two approaches are provided for Intrusion detection system: Boundary Cutting Algorithm and Clustering. Boundary Cutting Algorithm gives better accuracy than clustering as it is performs cutting according to space boundary.

Keywords*— Supervised learning, Unsupervised learning, Intrusion detection system, Data mining

I. INTRODUCTION

Now a day's internet is widely used. So securing computer network from threats and vernability is very important. Intrusion detection system helps for finding malicious activity. An Intrusion detection system gathers relevant data from computers or the network and analyses them for signs of intrusion. As shown in figure1 IDS is used as defensive mechanism in network. IDS are categories into three types [1]. (1) Host based IDS, (2) Network based IDS, (3) Distributed IDS. Host based IDS check a computer system to detect an intrusion. Network based IDS is used to analyse network capabilities. Network based IDS again divided into signature based IDS and anomaly based IDS. Anomaly based IDS are used to detect computer intrusion and monitor computer activity. After that it classified as normal or anomaly. Signature based IDS is mainly based on locking of known pattern. Distributed IDS is implemented for critical phases in the network.

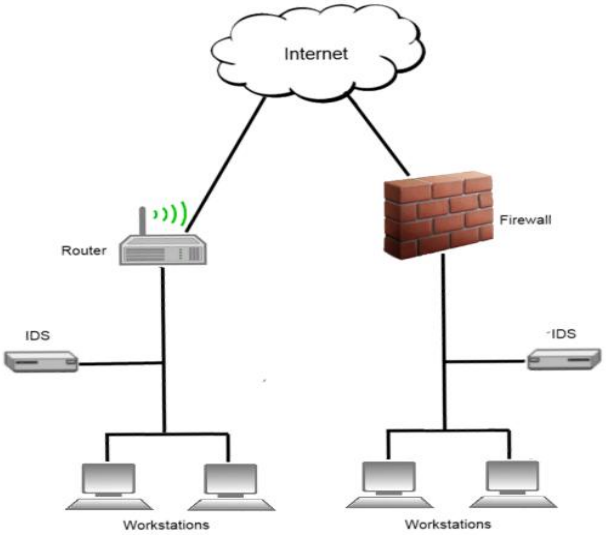

Figure1: Intrusion Detection System

Intrusion detection system algorithms can be classified into two types [2]: First is supervised learning approach and unsupervised learning approach. A supervised learning takes provided information and builds detection rule/model for known attacks. It has high detection rate and low false alarm rate. Unsupervised learning is able to learn new/ unknown attacks without training information. As compare with supervised learning it has lower detection rate and having high false alarm rate.

II. BACKGROUND

Different attacks are considered in IDS which requires large no of packets [13] as follow.

A. Denial of Service Attack

In Denial of Service attack, a machine or network resources are unavailable to intended users. It is divided into four parts: Attacker, handler, zombies and victim. Attacker is responsible for attacks and delivers command directly. These commands are received by handler. Handlers control all the zombies. These victims are attacked from different host.

B. Scanning Attack

Scanning attack is used by understanding all the information of system which being attacked. This technique allow any attacker to gain information about network like active host on network, which types of traffic firewall allowed, server software running, software version number, etc. Attackers use this information for launching of attacks.

C. Penetration attack

These types of attacks allow unauthorised attackers to access system and data. This happens because of exploiting software flow. These attack transport malicious payload to target host.

There are different attacks which are present in different protocols. In ICMP protocol, ICMP nuke attack**,** Ping of death, ICMP DOS Attacks are present. TCP is a connection oriented protocol. TCP protocol suffers from TCP SYN or TCP ACK Flood Attack, TCP Sequence Number Attack, TCP Hijacking and TCP reset attack. ARP flooding, User spoofing, Connection Hijacking and Interception attacks are present in ARP. UDP flood attack, Fraggle and Teardrop attack is in UDP.

Porras and Valdes [14] have proposed three parameters for evaluating the efficiency of IDS as follows:

1. Accuracy

Accuracy is the proper detection of attacks and the lower false alarms. Inaccuracy occurs when an intrusion-detection system flags a legitimate action in the environment as anomalous or intrusive.

2. Performance

The performance of IDS is calculated by using the rate at which audit events are processed. Real-time detection is not possible, if the performance of the intrusion-detection system is poor.

3. Completeness

Completeness is the ability of IDS to detect all attacks. Incompleteness occurs if intrusion detection system fails to detect an attack.

Data mining provide set of tools for IDS. Different data mining algorithm are used to implement Intrusion detection system such as support vector machine, decision tree, and Naive Bayes classifier and so on. In this paper we provide two algorithms for increasing detection accuracy. KDD CUP 1999 dataset is used for IDS.

III. RELATED WORK

In paper [5], authors proposed a data mining algorithm for intrusion detection system using ensemble classifier. They perform feature selection and multiboosting simultaneously. Binary classifier helps to improve the detection of attacks that occur less frequently in the training data. Based on this, new ensemble approach aggregates each binary classifier's decisions and decides which class is most suitable for a given input. Multiboosting is also used for reducing both variance and bias. This approach provides better performance and increase accuracy

Duanyang in [6] provides hybrid approach that combine host and network based ids to provide more effective intrusion detection system. This scheme is used for both anomaly and misuse detection.

In paper, Prashanth used random forest for network intrusion detection system [7]. This helps to increase accuracy. He considers approach for feature selection and optimization of parameters for random forest. In order to increase the rate of minority intrusion detection, the balanced dataset is constructed by using over-sampling the minority classes and down-sampling the majority classes. This balanced data set is used for building patterns using random forest which is much smaller than the original one.

In paper [8], Mehdi and Mohammad proposed a neural network approach for classification of attacks. Early shopping validation is used for increasing the generalization capability of the neural network and at the same time decreased the training time. The proposed system is able to classify records with 91% accuracy in case of hidden layers of neurons in the neural network and 87% accuracy with one hidden layer. Classification result is three layer networks.

Mrutyunjaya and Manas presented data mining based naïve Bayes classifier to classify whether attack is normal or anomaly [9]. In Bayesian classification, Hypothesis is used to determine that whether data belongs to particular class or not. Probability is calculated for hypothesis to be true. KDD cup'99 dataset is used for implementation. As compare to a back propagation neural network based approach, proposed technique performs better in terms of false positive rate, computational time and cost.

In paper [10], authors used classification model for misuse and anomaly attack detection using decision tree algorithm. C5, C4.5 and ID3 algorithms are used and comparison is done in order to analysed the performance IDS. C5 provides better performance.

In paper [11], authors presented multi-layer hybrid machine learning approach for IDS. In first layer, PCA is used for feature selection. Genetic algorithm generates anomaly detectors, which are able to recognize between normal and anomalous behaviours in the second layer. In order to increase the detection accuracy, classification is applied using naive Bayes, multilayer perceptron neural network, and decision trees.

Bhattacharyya presents clustering based classification method which is applied in Network intrusion detection system [12]. A set of labelled training data are divided into clusters based on similarity function. Clustering, training and prediction are provided. TUIDS Intrusion data set is used to evaluate clustering based classification method.TUDIS dataset consists of Packet level, Flow level and Port scan dataset.

Leila Mechtri et al., used principle component analysis for anomaly intrusion detection [5]. Experiment is done using the limited set of KDDcup99 and achieved 99% detection rate with 1% FP rate.

IV. PROPOSED METHOD

In order to detect intruders and increase the accuracy of detection, Two different algorithm is used.1) Boundary Cutting Algorithm and 2) clustering. These two algorithms are used for training and testing of incoming packets.
Figure1 shows the structure of Intrusion detection system. Before loading dataset there is need of pre-processing in order to remove inconsistent, redundant, and missing values. After that training is done using clustering and boundary cutting algorithm. According to that incoming packets are classified. And finally the accuracy of both algorithms is compare.

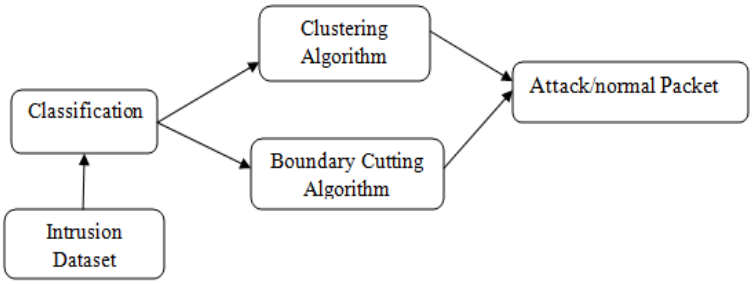

Figure 2.Architecture for IDS

A. Pre-processing

It is important step in data mining. Data may be consists of noise, inconsistent, incomplete and missing values. Pre-processing technique is used to clean such type of data. It includes normalization, transformation, feature extraction, etc.

B. Dataset

In this paper KDD CUP 1999 data set is used for intrusion detection system evolution competition. The data set consists of 41 attributes for the network traffic profiles with 4900000 data samples [4].

C. Classification

The goal of classification is to categorize data into distinct classes. Classification is two-step process. The first step is learning process. In this training data are analysed by a classifier algorithm. In second phase classification is done. Test data are used to estimate the accuracy of the classification rules. If accuracy is considered accepted, the rule can be applied to the classification of new data tuples.

D. Boundary Cutting Algorithm

Boundary Cutting Algorithm is an efficient packet classification algorithm. It finds out the space the each rule covers and performs cutting according to the space boundary [3].

E. Clustering

Clustering is nothing but grouping of objects. These objects are group together so that objects in one cluster are similar to each other than other clusters. Clustering is used in many fields like machine learning, pattern reorganization and information retrial.

V. CONCLUSIONS

In this paper, two approaches are given: Clustering and Boundary Cutting Algorithm. In testing phase, dataset is divided into clusters based on similarity. According to that clusters, classification is done to check whether that packet is normal or attack. Boundary cutting algorithm provides high accuracy than clustering as it perform cutting according to the space boundary.

ACKNOWLEDGMENT

I am thankful to my guide, HOD of computer Department and all teachers for their valuable guidance.

REFERENCES

[1] Rajendra Prasad Palnatya, Rajendra Prasad Palnaty, "JCADS: Semi-Supervised Clustering Algorithm for Network Anomaly Intrusion Detection Systems", IEEE, 2013.
[2] P. Jongsuebsuk, N. Wattanapongsakorn, C. Charnsripinyo "Network Intrusion Detection with Fuzzy Genetic Algorithm for Unknown Attacks", IEEE, ICOIN, 2013.
[3] Hyesook Lim, Nara Lee, Geumdan Jin, Jungwon Lee, Youngju Choi,Changhoon Yim, ,"Boundary Cutting for Packet Classification", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 22, NO. 2, APRIL 2014.
[4] Rajendra Prasad Palnaty, Prof. Ananda Rao Akepogu,"JCADS: Semi-Supervised Clustering Algorithm for Network Anomaly Intrusion Detection Systems", 2013 IEEE R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[5] Christine Dartigue, Hyun Ik Jang, Wenjun Zeng,"A New Data-Mining Based Approach for Network Intrusion Detection", in Seventh Annual Communication Networks and Services Research Conference, 2009.
[6] Duanyang Zhao, Qingxiang Xu, Zhilin Feng,"Analysis and Design for Intrusion Detection System Based on Data Mining", Second International Workshop on Education Technology and Computer Science, 2010.

[7] Prashanth, V.Prashanth, P.Jayashree, N.Srinivasan," Using Random Forests for Network-based Anomaly detection at Active routers", IEEE-International Conference on Signal processing, Communications and Networking Madras Institute of Technology, Anna University Chennai India, and Jan 4-6, 2008, Pp93-96.

[8] Mehdi MORADI and Mohammad ZULKERNINE, "A Neural Network Based System for Intrusion Detection and Classification of Attacks".

[9] Mrutyunjaya Panda and Manas Ranjan Patra ,"Network intrusion detection using naïve bays", in IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.12, December 2007.

[10] Manish Kumar, Dr. M. Hanumanthappa,"Intrusion Detection System Using Decision Tree Algorithm", IEEE, 2012.

[11] Amira Sayed A. Aziz Aboul Ella Hassanien Sanaa El-Ola Hanafy M.F.Tolba,"Multi-layer hybrid machine learning techniques for anomalies detection and classification approach", IEEE,2013.

[12] Prasanta Gogoi, B. Borah and D. K. Bhattacharyya,"Network Anomaly Identification using Supervised Classifier", Informatica 37 (2013) 93–105.

[13] Amrita Anand, Brajesh Patel "An Overview on Intrusion Detection System and Types of Attacks It Can Detect Considering Different Protocols", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 8, August 2012.

[14] Phillip A. Porras and Alfonso Valdes. Live traffic analysis of TCP/IP gateways. In Proceedings of the 1998 ISOC Symposium on Network and Distributed System Security (NDSS'98), San Diego, CA, March1998. Internet Society.

