ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management (IJIRTM), Volume-3, Issue-4, 2019

NB Tree based Intrusion Detection Technique using Rough Set Theory Model

Vivek Kumar[1], Prof. Narendra Parmar[2]

[1]M. Tech Scholar, Department of CSE, SSSCE, Bhopal (India)

[2]Assistant Professor, Department of CSE, SSSCE, Bhopal (India)

ABSTRACT

Intrusion detection System is a defense measure that supervises activities of the computer network and reports the malicious activities to the network administrator. Intruders do many attempts to gain access to the network and try to harm the organization's data. Thus the security is the most important aspect for any type of organization. Due to these reasons, intrusion detection has been an important research issue. Rough sets are the sets defined through these upper and lower approximations. Rough set theory has turned into an important and sophisticated device in the determination of a wide range of various issues such as showing uncertainty or general knowledge, knowledge discovery, estimation of the quality and availability of information with respect to the existence of a note of date patterns, evaluation and identification of date dependency with the hybrid approach used improve accuracy and f-measure.

Keywords:- Rough set theory, tree based classifier, signature-based IDS, anomaly-based IDS.

INTRODUCTION

An intruder is a person who tries to gain unauthorized access to a system, so that the system can be damaged and data on that system can be disturbed. Intrusions are usually caused by intruder's attackers, who want unauthorized and additional privileges to particular system or network for their own purposes the number and severity of the network attacks have increased in past few years. So for securing the network from different network attacks, intrusion detection system (IDS) plays a key role [2]. Detection of the intrusive activities by using resource intensive intelligent algorithms has been possible because of advancements in computing performance in terms of processing power and storage. Intrusion detection can be explained as the act of detecting actions that attempt to compromise the confidentiality, integrity or availability of a resource. The main goal of intrusion detection is to identify entities attempting to subvert in-place security controls. A system called intrusion detection is way for detecting abnormal behaviors in a system. An unexpected pattern covers many definitions but in general it is likely represent as undesirable, malignant and/or misuse activity occurring within a system.

The two general way described are two different capability of spotting intrusions. As mentioned intrusion detections can be deployed on different areas, comparable within a computer to spot users attempting to gain access to which they have no access right, or monitoring network traffic to detect other kind of intrusions like worms, trojan horses or to take control of a host by yielding an illegal root shell. Over the previous decade, web clients have seen an exponential development in the quantity of site pages available through well known web crawlers. Sorting out the vast volume of web data in an all around requested and exact way is basic for utilizing it as a data asset. Site

page arrangement addresses the issue of doling out predefined classifications to the website pages by method for directed learning. This inductive learning process naturally fabricates a model over an arrangement of already characterized site pages. The scholarly model is then used to order new website pages. Various classifiers proposed and utilized for machine learning can be connected for site page arrangement. These incorporate support vector machines (SVMs), k-nearest neighbor (k-NN), and naive bayes (NB) classifiers.

1. Basic concepts of Intrusion detection system (IDS)
IDS are usually deployed along with some other preventive security controls mechanism, such as access control and authentication, as a second line of defense which protects the information systems. There are two main techniques of intrusion detection misuse & anomaly detection. Anomaly detection detects unusual activity patters in the observed data. It is based on subject's (e.g. a system or a user) normal behavior. Any significant deviation from the usual activity is considered as intrusive. Misuse detection technique recognizes known attack patterns. It is based on signature of known attack. Any action that matches with the pattern of a known attack is considered as intrusive.

2. Neural network
It is one of the most up to date flags preparing innovation. ANN is a versatile, nonlinear framework out how to play out a capacity from information and that versatile stage is regularly preparing stage where framework parameter is change amid operations. After the preparation is finished the parameter are settled. On the off chance that there are loads of information and issue is inadequately reasonable then utilizing ANN model is precise, the non direct attributes of ANN give it bunches of adaptability to accomplish input yield outline. Fake neural networks, give client the capacities to choose the system topology, execution parameter, learning guideline and ceasing criteria.

3. Rough set theory
Rough set theory (RST) has emerged as an intelligent tool for knowledge discovery from imprecise, uncertain datasets through the identification of redacts (feature subset) which represents the maximum information of the system [3]. The rough set theory is an important mathematical tool to deal with imprecise, inconsistent, incomplete information and knowledge [4]. Originated from the simple information model, the basic idea of the rough set theory can be divided into two parts.

(i) The first part is to form concepts and rules through the classification of relational database.
(ii) The second part is discovery knowledge through the classification of the equivalence relation and classification for the approximation of the target.

NB tree
Bayes networks are one of the most widely used graphical models to represent and handle ambiguous information. Bayes networks are specified by two components:

(i) A graphical component composed of a directed acyclic graph (DAG) where vertices represent event and edges are relations between events.

(ii) A numerical component consisting in a evaluation of different links in the DAG by a conditional probability distribution of each node in the context of its parents.

Naive bayes are very simple bayes networks which are possessed of DAGs with only one root node (called parent),representing the unobserved node, and several children, comparable to observed nodes, with the strong assumption of independence among child nodes in the context of their parent. The classification is ensured by considering the parent node to be a hidden variable stating to which class each object in the testing set should belong and child nodes represent different attributes specifying this object.

II LITERATURE REVIEW

Literature review for different concept and techniques are describe below

Zhang et al. the rough set theory has been researched for more than thirty years and it has made many achievements in many fields, such as machine learning, knowledge acquisition, decision analysis, knowledge discovery in database, expert system, decision support system, inductive inference, conflict resolution, pattern recognition, fuzzy control, medical diagnostics applications and research on granular computing, it has become one of the main models and tools. The application prospect of rough set theory is very broad. The rough sets not only can be used to deal with new uncertain information systems, but also can optimize many existing soft computing methods. For the rough set theory, in the process of data mining, there are still a large number of problems need to be discussed, such as large data sets, efficient reduction algorithm, parallel computing, hybrid algorithm, etc.

Paul et al. they have presented that expansion in the measure of data on the attack has brought on the requirement for exact mechanized classifiers for Attack pages to keep up attack registries and to expand internet searcher execution. Each tag and each term on every attack page can be considered as an element there is a requirement for productive techniques to choose best components to lessen includes space of the attack page order issue. Apply a late enhancement method specifically the CART to choose best elements for attack page characterization issue. The CART is a metaheuristic calculation, enlivened by the blazing conduct of fireflies. Utilizing FA to choose a subset of elements and to assess the wellness of the chose highlights J48 classifier of the weka information mining instrument is utilized. Attack KB and conference datasets were utilized to assess the viability of the proposed highlight determination framework. Perception is that when a subset of elements are chosen by utilizing FA, attack KB and conference datasets were grouped without loss of precision much additional time expected to order new attack pages diminished forcefully as the quantity of elements were diminished.

Thaseen et al. it is aimed at evaluating different tree based classification algorithms that classify network events in intrusion detection systems.

Experiments are conducted on NSL-KDD 99 dataset. Dimensionality of the attribute of the dataset is reduced. The results show that random tree model holds the highest degree of accuracy and reduced false alarm rate. Random tree model is evaluated with other leading intrusion detection models to determine its better predictive accuracy.

Ingre et al. intrusion detection system based on decision tree has been for NSL-KDD dataset. Four different models have been proposed for the dataset two model for five class classification (normal and types of attack) in which the model uses feature selection. Likewise a different model is created for binary class classification (normal and attack). The system uses CART algorithm with gini index as splitting criteria for pattern classification and correlation based feature selection (CFS) is used for dimensionality reduction.

III PROBLEM STATEMENT

Literature review several techniques have been proposed to improve the efficiency of classification. User required searching his desire CART techniques through the search engines. There is still a problem of mapping and detection of the tree, to work effectively. Some time they give good results whereas most of the time the result is anonymous because it is required to work on some of the parameters like the accuracy and F-measure

IV OBJECTIVE

Various works has been done under this a device or software which monitors network traffic and suspicious activity, if any deviation occurs against normal behavior, then it alerts the system or network administrator, as on the following parameters:

(i) Accuracy

(ii) F-Measure.

ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management (IJIRTM), Volume-3, Issue-4, 2019

V RESULT ANALYSIS

A. Experiment to calculate accuracy between CART and NB-IDS

In the experiments both the CART and NB-IDS are trained using Kyoto dataset. This experiment is carried to evaluate accuracy of the CART and NB-IDS. Experiments are performed with variation in dataset from 10% to 50%. Increasing value of dataset indicates how much value of self data is available for training

Table 1: Accuracy of CART and NB-IDS

Dataset %	CART	NB-IDS
10	0.87	0.93
20	0.88	0.94
30	0.88	0.95
40	0.89	0.95
50	0.9	0.96

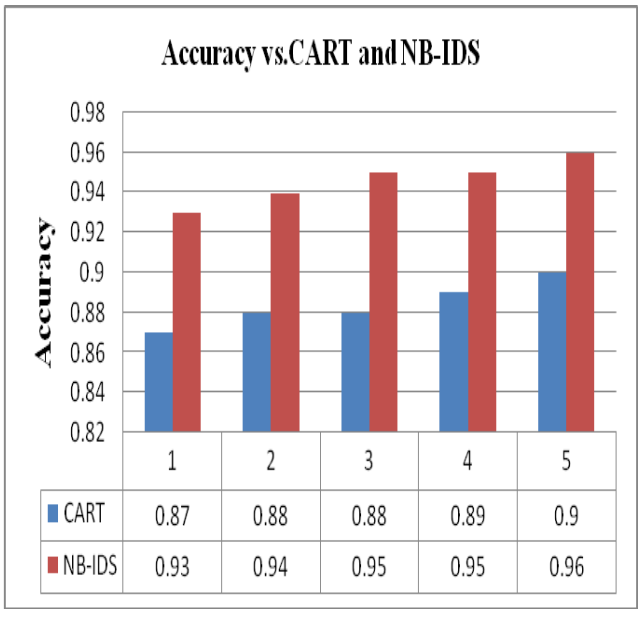

Figure 1: Accuracy comparison between CART and NB-IDS.

B. Experiment to calculate F-measure between CART and NB-IDS

Experiments are performed with variation in dataset from 10% to 50%. Experiments are carried with increasing value of dataset that indicates how much value of self data is available for training

Table 2: F-measure of CART and NB-IDS

Dataset %	CART	NB-IDS
10	0.89	0.93
20	0.86	0.94
30	0.84	0.95
40	0.8	0.95
50	0.9	0.96

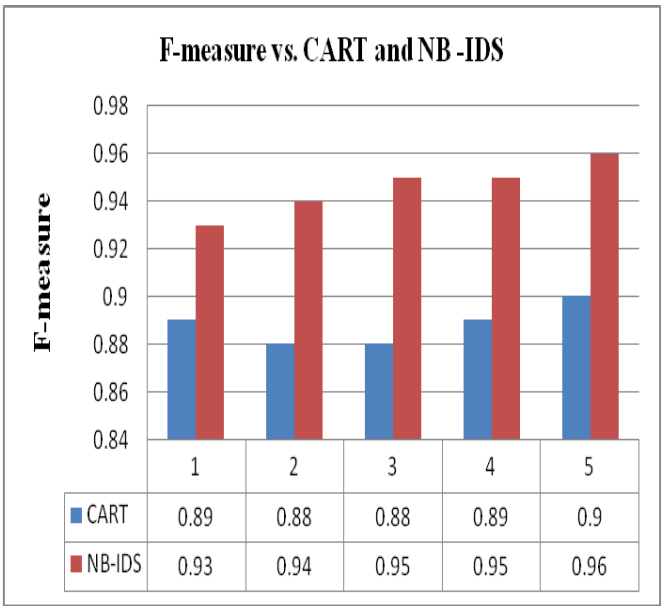

Figure 2: F-measure comparison between CART and NB-IDS.

VI CONCLUSION

The performance of the dissertation work is compared with the performance of various techniques.

- The comparison of their performance is check based on the accuracy in detecting attacks on the test of kyoto dataset.
- Intrusion detection system can be categorized as a security tool which inspects any inbound and outbound network intrusion activities against the PC or systems such as illegal access, misuses and any type of online hacker attacks, conventional security technology such as user authentication, data security, and firewalls which have been used as the first layer of computer protection, do not assure the security of system absolutely.
- This dissertation outlined the rough set theory and tree-based classifier in IDS which outperformed the existing methods substantially.
- This dissertation has been talks about the efficiency of the proposed work over existing work in section.

Clearly shows that the proposed work performs much better than existing work on the parameter of the accuracy and F-measure.

REFERENCE:-

1. Deshpande,V.K.: Intrusion detection system using decision tree based attribute weighted AODE, International Journal of Advanced Research in Computer and Communication engineering,pp.8738-8743,(2014).
2. Rai, K.M., Devi,S., Panjab,G., Chekuri,v.: Decision tree based algorithm for intrusion detection, International Journal of advanced networking and applications, pp.2828-2834, (2016).
3. Raman,M.R.G., Kannan,K., Pal,S.K. and Shankar V.: Rough Set-hyper graph-based feature selection approach for intrusion detection systems, Defence Science Journal,pp.612-617, (2016).
4. Pawlak, Z.: Rough set theory, International Journal of. Computer science, pp.7-10, (1982).
5. Li,T., Nguyen,H.S., Wang,G.Y.: Rough Sets and Knowledge Technology Springer Berlin Heidelberg (2012).
6. Zhang,Q.,Xie,Q., Wang,G.: “A survey on rough set theory and its applications”, CAAI transactions on intelligence technology,pp.323-333,(2016).
7. Paul,P., Dey,U., Roy,P., Goswami,S.: “Network Security Based on Rough Set theory, pp. 165-169, (2015).
8. Thaseen,S., Kumar,A.,: “An analysis of supervised tree based classifiers for intrusion detection system”, Pattern Recognition, Informatics and Mobile Engineering (PRIME), (2013).
9. Ingre,B., Yadav,A., Soni,A.K.: Decision tree based intrusion detection system for NSL-KDD Dataset, (2017).
10. Bordbar,S., Chakrabarti,k.: Abdulah M.b.: A feature selection based on rough set for improving intrusion detection system, (2015).
11. Saurabh,P.,Verma,B, Sharma,S.: Biologically Inspired Computer Security System: The Way Ahead, Recent Trends in Computer Networks and Distributed Systems Security, Springer, pp 474-484, (2011).
12. Saurabh,P.,Verma,B,: Cooperative Negative Selection Algorithm, International Journal of Computer Applications (0975 – 8887), vol 95 - Number 17, pp 27-32, (2014).
13. Saurabh,P.,Verma,B, Sharma,S.: An Immunity Inspired Anomaly Detection System: A General Framework A General Framework, Proceedings of 7th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012), Springer, pp 417-428, (2012).
14. Saurabh,P.,Verma,B.: An Efficient Proactive Artificial Immune System based Anomaly Detection and Prevention System, Expert Systems With Applications, Elsevier, 60, pp 311–320, (2016).