International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-6 Issue-2, December 2016

Intrusion Detection over Networking KDD Dataset using Enhance Mining Algorithm

Bhagwat P. Dwivedi, Shiv Kumar, Babita Pathik

Abstract: ***The intrusion detection systems (IDSs) generate large number of alarms most of which are false positives. Fortunately, there are reasons for triggering alarms where most of these reasons are not attacks. In this research, a rule based technique which is the enhancement of genetic algorithm has been developed. For this, The networking data and intrusion over the data is find to extract to recognize various entities into it. Data mining and its algorithm to process, data extraction, and data analysis is an important phase to monitor the features in it. Intrusion detection process follows the clustering and classification technique to monitor the data flow in it. In this paper our investigation is about to observe available algorithm for the intrusion detection. Algorithm such as Genetic, SVM etc have been processed over KDD cup 10% of dataset which contain 41 attributes and large number of data availability. Here our experiment also conclude that the proposed feature extraction algorithm outperform as best than the existing algorithm with computation parameter such as precision, recall and its accuracy.***

Keywords: ***Intrusion detection, clustering technique, Data mining, KDD.***

I. INTRODUCTION

Classification, regression and clustering are three approaches of data mining in which instances are grouped into identified classes. Classification is a popular task in data mining especially in knowledge discovery and future plan. It provides the intelligent decision making. Classification not only studies and examines the existing sample data but also predicts the future behavior of that sample data. It maps the data into the predefined class and groups. It is used to predict group membership for data instances. In Classification, the problem includes two phases first is the learning process phase in which for analysis of training data, the rule and pattern are created.

The number of hacking and intrusions incidents is increasing year on year as technology rolls out. Maintaining a high level security to ensure safe and trusted communication of information between various organizations becomes a major issue. So Intrusion detection system (IDS) has become a needful component in terms of computer and network security [1].

Manuscript published on 30 December 2016.
***** Correspondence Author (s)
Bhagwat P. Dwivedi, M.Tech. Scholar, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.
Dr. Shiv Kumar, Professor & Head, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.
Babita Pathik, Assistant Professor, Department of Computer Science and Engineering, Lakshmi Narain College of Technology Excellence, Bhopal (M.P)-462021, India.

An Intrusion Detection system (IDS) is a device or a software product that analyzes the coming traffic on network for a malicious activities (or intrusion) and raises an alarm when intrusion detected. The aim of IDS is to detect illegal and improper use of system resources by unauthorized users by monitoring network traffic and audit data. An intrusion can be defined as any set of actions that attempt to compromise the integrity, confidentiality or availability of resources on system [2] [3].

II. RELATED WORK

Today in Dataset there exist data objects [4] that do not comply with the general behavior or model of the data. Such data objects, which are heavy different from or inconsistent with the remaining set of data, are called outliers. An outlier is a data set which is different from the remaining data. Outlier is also referred to as deformity, deviants or anomalies in the data mining and statistics literature. In most applications the data is created by one or more generating processes, which could either reflect activity in the system or observations collected about entities. When the developing process behaves in a casual way, it results in the creation of outliers. Therefore, an outlier [5] often contains useful information about anomaly characteristics of the systems and entities, which impact the data generation process. The recognition of such unusual characteristics provides useful application specific insights.

In the base paper [9] it discussed about the genetic based feature selection algorithm where they have taken full NSL Kdd[10] dataset and perform the intrusion detection system over it. They have used genetic algorithm for searching for the feature selection and getting outlier using the technique. Further the parameter taken and shown by them as computation time and growth in accuracy. They have compared their work with Naïve based with more reduce feature. [1]

The techniques which are been performed for the outlier detection and in order to work with the Intrusion detection system with the data set provided. We have taken the various paper of different latest author and understood the work performed by them and finally we have concluded the discussion related to the previously performed algorithm in IDS system [6]. Upon discussion of previous algorithm we are able to understand the further work which can be performed for outlier detection in KDD dataset.

III. PROBLEM FORMULATION & PROPOSED WORK

As per the observation and discussion on previous algorithm with the intrusion detection and data mining approach over the kdd cup dataset. Here are the problem formulation faced with the current scenario availability [7].

1. The existing technique based on the previous rules defined and few rules are in repetitive process.
2. Existing technique perform the other detection also along with the intrusion, which often make wrong use of data and experiment.
3. The existing approach perform [8] the complete process of parsing, noise removal and then performance. Such that a long interval and process need to involve in it.

A proposed rule based algorithm which is the enhancement of genetic approach take participate in communication such that an optimized rule can be perform and proper outlier using the technique can opt out. The further approach follow the rule extraction , mutation over the data processing, further crossover and then rule applying over the process. Thus the better outcome observation can be taken.

Algorithm Pseudo Code:

Input: Dataset Kddcup99, Rule parameter, conditional parameter.
Output: Outliers, intrusion related data.
Steps: Begin – Load all Data(Kddi-n)
foreach{
finding clusters();
rolling over complete ds();
read(Kdd attribute features);
}Performing ruleover
{
Mutation();
EnhanceFunc();
Crossover();
Rule applying over refined data();
}
Performing Parameter extraction()
{
Compute TP, FP, TN, FN;
Compute the parameter using formulae;
Return parameter values;
}
End;

The above complete process determines the work performed and processed by us. Thus an effective sapproach and enhance genetic rule over the algorithm is performed which is compared with traditional genetic approach and SVM approach[4]. The further result and other discussion is made on result and experiment basis.

IV. EXPERIMENTAL SETUP & RESULT ANALYSIS

In order to perform experiment over the dataset and technique with the software's and tools. Java Language with its working IDE net beans 8.0 is taken for the experiment framework designed using Swing API. The workbench setup is done using the dataset KDD cup 99, 10% of the dataset is taken for the experiment containing 10 main attributes for the work. All the experiment performed using core i5 CPU with 4 GB of RAM to observe the data processing and its performance.

Dataset: Data set which is KDD cup 1998, 10% of data is taken for experiment. Performance Measures To estimate the performance of the system, the following formulas are used.

Classification rate = (Number of classified patterns * 100)/ Total number of patterns.

True positive rate = TP/TP+FN

False positive rate = FP/FP+TN

On applying probability based feature selection algorithm, following features were selected.

- There are result parameter to monitor the efficiency of that algorithm is as follows[5,6]-
- Precision – this parameter value can be calculated by processing the algorithm, precision is number of retrieval data to the query..
- Precision = (relevant data ^ retrieved data)/retrieved data.
- Recall – this parameter value can be calculated by processing the algorithm, Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved.

Recall = (relevant data ^ retrieved data)/relevant data

As per the observed result and experiment setup[7], technique is implemented. The proposed and existing technique is performed with the above dataset presented and the algorithm performed with the system and following output results was monitored.

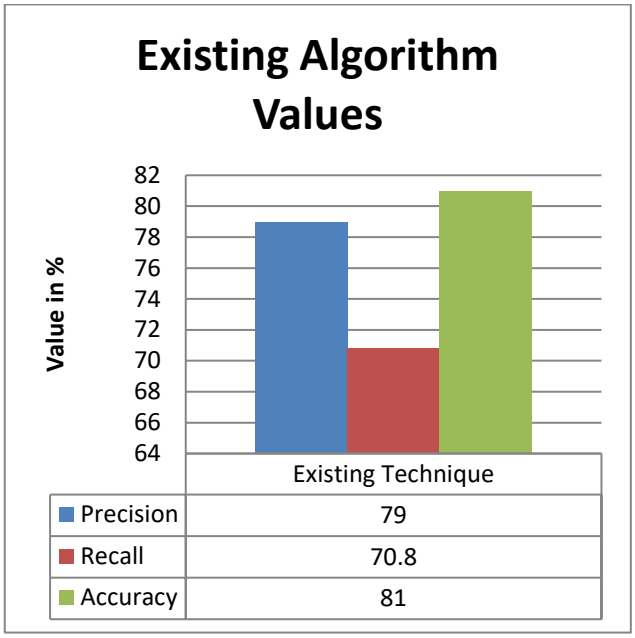
In the table present below is a statistical comparison is performed.

Table 1: Data Result Comparison.

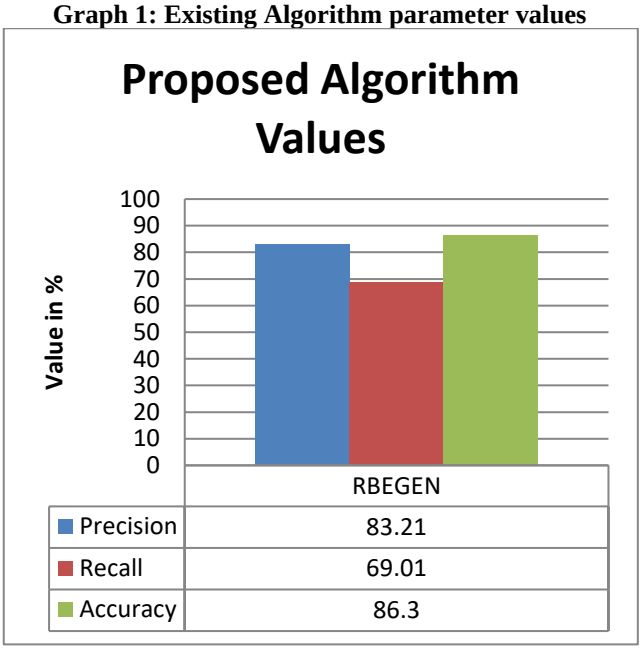
Technique Approach	Existing Technique Model	Proposed RBEGEN Algorithm
Precision	79	83.21
Recall	70.8	69.01
Accuracy	81	86.30

The above table represents the number of data values from the data and algorithm is performed.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

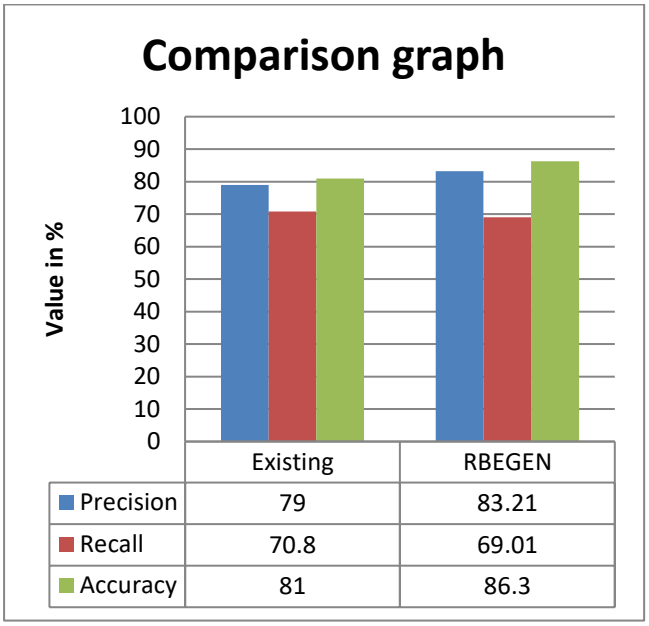


Graph 1: Existing Algorithm parameter values



Graph 2: Proposed approach parameters

The above graph no.1& graph no.2 represent the number of data values from the data and algorithm is performed.



Graph 3: Comparison graph for technique analysis x axis in % parameter value.

In the above graph drawn x axis as data from which technique were taken for the image data processing for specified dataset and bar graph is printed using the chart library provided by the Microsoft and further analysis can easily performed thus the RBEGEN approach outperform the best The graph representation shows the efficiency of that proposed algorithm work and it outperforms effective parameter value[8].

V. CONCLUSION AND FUTURE WORK

The Intrusion and outlier detection algorithm are the important factor to discuss which can resolve and determine using data mining algorithms. There are few previous algorithms such as SVM, Naïve bayes, Genetic based and other rule based approach for detection. In the paper a proposed Rule based Enhance genetic (RBEGEN) approach is followed, where the new rules on parameters been applied such that an high accuracy and other parameter can obtain. Our experiment gives the extensive result compare to existing algorithm with same dataset. Thus the approach can outperform at best with existing approach. Our further work will be in dealing with its real time application such as antivirus and other instruction application etc.

REFERENCE

1. Zhan Jiuhua Intrusion Detection System Based on Data Mining Knowledge Discovery and Data Mining, 2008. WKDD 2008.
2. Bane Raman Raghunath Network Intrusion Detection System (NIDS)Emerging Trends in Engineering and Technology, 2008. ICETET '08.
3. Changxin Song Design of Intrusion Detection System Based on Data Mining Algorithm 2009 International Conference on Signal Processing Systems.
4. Wang Pu Intrusion detection system with the data mining technologies Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

5. Gaikwad, D.P. Sonali Jagtap, Kunal Thakare, Vaishali Budhawant Anomaly Based Intrusion Detection System Using Artificial Neural Network and fuzzy clustering International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, 1 (9.) (2012 November).
6. Goyal, C. Kumar GA-NIDS: A Genetic Algorithm based Network Intrusion Detection System, Electrical Engineering and Computer Science, North West University Technical Report (2008).
7. G. Gu, P. Porras, V. Yegneswaran, M. Fong, W. Lee BotHunter: detecting malware infection through IDS-driven ialog correlation Proc. of 16th USENIX Security Symp. (SS'07) (2007 Aug), pp. 12:1–12:16.
8. G. Gu, J. Zhang, W. Lee BotSniffer: detecting botnet command and control channels in network traffic Proc. of 15th Ann. Network and Distributed Sytem Security Symp. (NDSS'08) (2008 Feb).
9. Ketan Sanjay Desale, Roshani Ade," Genetic algorithm based feature selection approach for effective intrusion detection system", IEEE 2015.
10. Kajal rai, "Decision Tree Based Algorithm for Intrusion Detection", Volume: 07 Issue: 04 Pages: 2828-2834 (2016) ISSN: 0975-0290.