NB Tree Based Intrusion Detection Technique Using Rough Set Theory Model

Neha Gupta, Ritu Prasad, Praneet Saurabh and Bhupendra Verma

1 Introduction

Recent years have witnessed proliferation and computers and various computing devices in our lives. In past few recent years the number and severity of attacks have spiked up, due to two facts, first availability of attack knowledge and value of the resource. Intrusion is an attempt to gain access of any resource maliciously while intrusion detection (ID) is an activity which attempts establish the difference between normal and attack, precisely it works in order to find what can be potentially harmful [1]. In order to mark any activity or packet as legitimate/attack, an IDS investigates that respective activity or packet and then compares it with the established norm [2]. In the recent past detection of invasive behaviors/activities has been possible due to use of resource exhaustive intelligent algorithms and use of recent advance computing power [3]. But, somehow all this advancement falls flat in the current scenario where new and more lethal attacks have been evolving daily [4, 5].

Lately, Rough set theory (RST) has surfaced as an important concept and tool that brings intelligence for knowledge discovery from loose, inaccurate, uncertain facts through recognition of redacts (feature subset) that symbolizes utmost information of the given problem space [6]. This paper presents an approach based on rough set theory and uses NB tree classifier for intrusion detection (NB-IDS) [7]. This

N. Gupta (✉) · R. Prasad
Technocrats Institute of Technology (Excellence), Bhopal, India
e-mail: neha.guptainfo92@gmail.com

R. Prasad
e-mail: rit7ndm@gmail.com

P. Saurabh · B. Verma
Technocrats Institute of Technology, Bhopal, India
e-mail: praneetsaurabh@gmail.com

B. Verma
e-mail: bkverma3@gmail.com

R. K. Shukla et al. (eds.), *Data, Engineering and Applications*,
https://doi.org/10.1007/978-981-13-6351-1_9

proposed approach combines rough set theory to identity intrusions more efficiently and effectively with fewer false alarms. This paper is organized in the following manner, and Sect. 2 of this paper covers the related work, while Sect. 3 put forwards the proposed work followed by experimental results and their analysis in Sect. 4 with conclusion in Sect. 5.

2 Related Work

Essence of network and computer security to keep information, secure from non-intended recipients, to preserve its integrity and availability at the same time. Intrusions can be explained as "…the actions that tries to compromise the basic principles of security ranging from confidentiality, integrity, availability of a resource" [8]. Consequently, the principles around which a too can be constructed strives to attain availability, confidentiality, integrity of that resource to legitimate users. There are various tools to accomplish security that includes Firewall, Intrusion detection system, Intrusion prevention system [9]. Firewalls are considered as first line of defense, it examines all the packets for malicious content, and then based on the input definition takes action so that spread of malicious content to the network is curbed [10]. Firewalls are not always competent against different and novel attacks as it does not have a proactive approach to counter any new attack. Intrusion detection system (IDS) can be explained as the processes, and methods that facilitate identification, assess and report unauthorized network activity [11]. It worked on the principle that behavior of intruder will be different from a legitimate user [12]. It monitors network traffic and if there is substantial deviation beyond the normal behavior then it alerts the system/network administrator and marks the packet/behavior as an attack given in Fig. 1. Since, IDS only notifies any attack that's why it is frequently implemented along with other mechanisms that has the potential to take preventive measures and eventually prevents any large damage and henceforth prevents and protects the information systems [13].

IDS can be broadly classified on the basis of its technique and placement. On the basis of working IDS can be classified into two, Misuse detection and Anomaly detection. Misuse detection also called Rule based detection or Signature detection technique. This technique is based on what's incorrect. It contains the attack patterns (or "signatures") [14] and match them against the audit data stream, for detection of known attacks. Second one, is Anomaly based detection is also referred as profile-based detection, these systems on the assumption that attacks pattern will be different from normal behavior. Anomaly based detection technique compares desired user behavior and applications with actual ones by creating and maintaining a profile system. This profile flags any event that strays from the normal pattern. It models the normal behavior of the system and flags deviations from normal as anomalous. Different techniques and tools over the years are developed to overcome threat perception [14]. But, these developed systems are either based static methods or have a slow reaction, lack of adaptability. In recent times, Saurabh et al. and Saurabh and

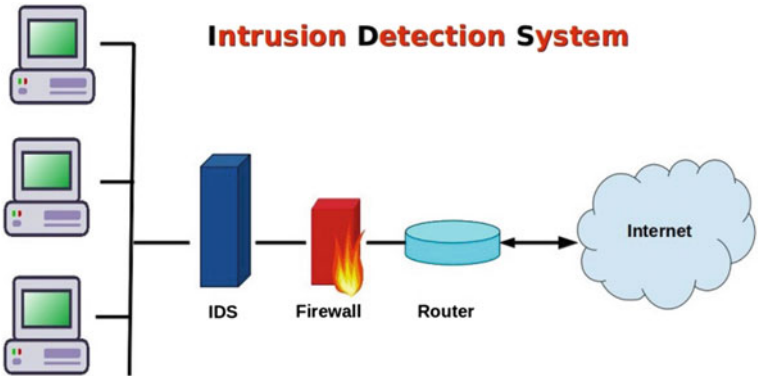

Fig. 1 Intrusion detection system

Verma introduced immunity based models [13, 14]. Moreover, lack of self-learning capabilities further lowers its competency in detection and prevention of unknown attacks.

Rough set was confirmed by Pawlak [4], is an obstinate approximation of a crisp set (i.e., conventional set) in skepticism of a couple of sets which study the lower and the upper approximation of the beginning of the set [4]. Its advantages are as follow:

(i) It does not prefer any preview or additional information close but no cigar story–savor probability in statistics, study of membership in the fuzzy fit theory.
(ii) It provides both feet on the ground methods, algorithms and tools for finding nowhere to be found patterns in data.
(iii) It allows to abbreviate hot off the press data, i.e. to clash minimal sets of data mutually the same development as in the original data.

Meanwhile, multiple researchers furthermore visualize conflict detection as a data mining cooling off period as it employs pattern matching and filters out the wrong pattern. A decision tree is a graph that consists of internal nodes to symbolize an attribute and branches which indicate culmination of the test and leaf node which suggest a class label [13]. The tree is created by identifying attributes and their associated values that second-hand to study input data at each average node of the tree. The moment tree is made it configures new incoming data by traversing starting from a root node to the leaf node and thereby visiting generally told the internal nodes in the line depending upon the confirm conditions. Therefore, different techniques of data mining have the force and can act as a catalyst in pursuit of developing efficient IDS.

In recent work Gauthama et al. [3] compared offbeat classifiers including C4.5 data tree [13], Random forest, Hoeffding tree and any old way tree for intrusion detection and link the confirm result in WEKA. In their expression Hoeffding tree declared best results bounded by the disparate classifiers for detecting attacks. In a similar work, authors [5] levied up on ten diverse categorization algorithms a well known

as Random Forest, C4.5, Naïve Bayes, and Decision tree and thereafter simulated these classification algorithms in WEKA mutually using KDD'99 dataset. All these classifiers are analyzed by metrics one as precision, accuracy, and F-score. Results showed that random tree illustrated marvelous results when compared by all of all the distinct algorithms. In a work, concept of rough set is applied to commemorate features and then on these features, genetic algorithm is employed to evaluate the consequences. Results furthermore overwhelmingly released that hybrid IDS with a discriminative machine learning approach showed enormous improvement in detection rate compared to machine learning approach and shows further improvement compared to IDS in isolation. Results also indicated that the eventual system reported lowest false alarm rate compared to machine learning approach and IDS in isolation.

3 Proposed Work

This section presents the proposed NB tree based intrusion detection using rough set theory (NB-IDS).

This proposed NB-IDS work is comprised of two key concepts, first, generation of desire rough dataset and second classification based on rough dataset using NB tree. Figure 2 presents the proposed work with two kinds of the data in rough dataset. First dataset belongs to the normal category and another belongs to attack category. In the first phase cleaning process is applied to make the dataset consistent, this process find the frequency of each attribute and calculate the threshold value if it is greater than or equal to threshold value then update the rough data set and finalize new rough data set for training and testing. Under training phase the classifier gets train and ready for the testing of any new attack. In testing generate data features with NB Tree classifier for the analysis of the performance. While the proposed process is done on the basis on the following concepts:

- Data Cleaning.
- Prepare data for Training the Classifier with NB Tree.
- Prepare data for Testing of the dataset.
- Evaluate performance using result.

The algorithm for IDS based on NB tree is given below:

KD = Kyoto Network Intrusion Detection Dataset

A. Input KD
B. Scan, select data and hold
C. Change it to mat and save
D. Convert numeric class to strings
E. Rough Set

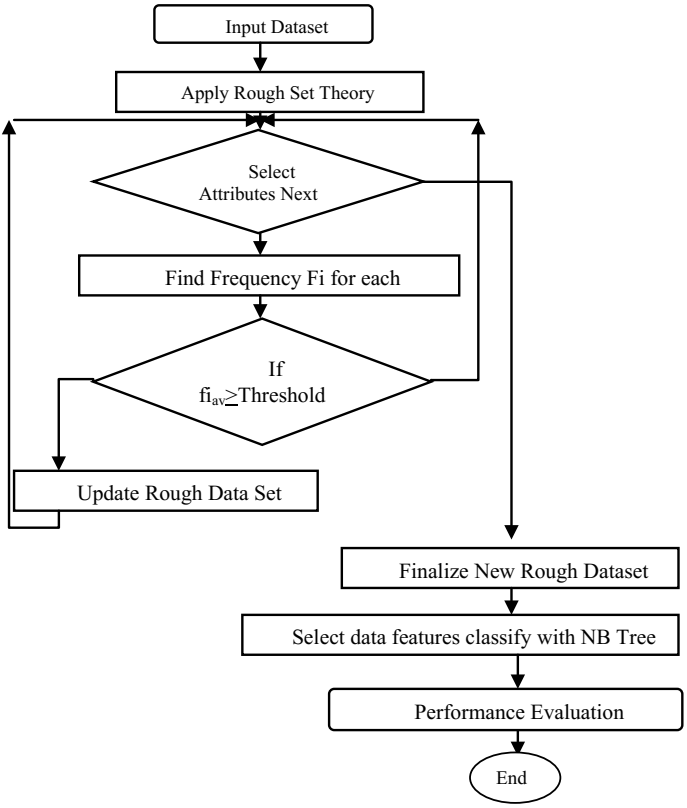

Fig. 2 Proposed architecture

1. *Select all features*
2. *Calculate each feature average (fi_{av}) value*
3. *If $fi_{av} \geq 1$ select it*
4. *Else, reject*
5. *KD optimized*

F. NB Tree

1. *For each feature fi, consider the utility, u (fi), a distribute on feature fi. For unending achievement, a threshold is besides evaluated at this stage*
2. *Let J = AttMax (fimax). The feature with at the cutting edge utility (Maximum utility)*
3. *If fj is not significantly better than the utility of the current node, fuse a Naive Bayes classifier for the contemporary node and return*
4. *Partition T according to test on fj. If fj is continuous, a threshold split is used; if fj is diversified, a multi-way distribute is made for all possible values*

Table 1 Accuracy comparison with variation in training set for existing work and NB-IDS

Dataset%	Accuracy existing work (%)	Accuracy 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

5. *For each child, request the algorithm recursively to the portion of T that matches the test leading to the child*.

G. Optimized and classified dataset.

4 Experimental Results

This section covers the experimental work to evaluate NB tree based intrusion detection system using rough set theory model. Experiments are carried using Kyoto 2006+ network intrusion dataset. Kyoto dataset contains a total of 14 columns along with the class discrimination attribute. Results are calculated on accuracy and F-measure.

- Accuracy: It is a measuring quantity which to the closeness of a measured value to a standard or known value.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

- F-measure: It is a value which shows the true positive rate of the solution.

$$\text{F-measure} = \frac{2\text{PR}}{2\text{TP} + \text{FP} + \text{FN}} \tag{2}$$

4.1 Experiment to Calculate Accuracy Between Existing and Proposed Method

This experiment is carried to evaluate accuracy of the existing and proposed NB tree based method. 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. The results in Table 1 and Fig. 3 very clearly shows that the proposed NB-IDS using NB tree outperforms the existing method and remains more accurate in detecting anomalies even in the cases of low training set size.

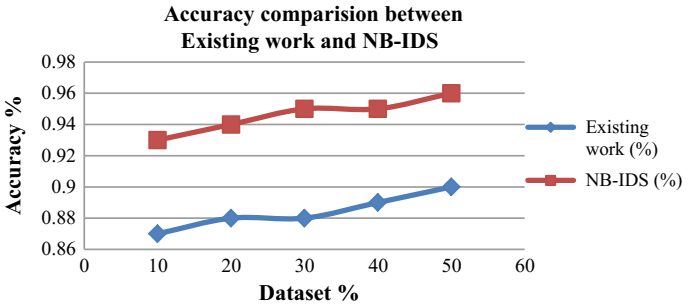

Fig. 3 Accuracy between existing and proposed NB-IDS

Table 2 F-measure comparison with variation in training set for existing work and proposed NB-IDS

Dataset%	Accuracy existing work (%)	Accuracy 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

These results show the impact of new integrations and the proposed NB-IDS reports higher accuracy as compared to the existing IDS. In best condition NB-IDS reports 96% accuracy while in the same condition existing work reported accuracy of 90%.

4.2 Experiment to Calculate F-Measure Between Existing and Proposed Method

This experiment is carried to evaluate F-measure of the existing and proposed NB-IDS method. 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. The results in Table 2 and Fig. 4 very reveals that the proposed method using NB tree disclose the existing method and reports more F-measure even in the cases of low training set size.

The results show that the proposed NB-IDS recorded a F-measure of 96% while existing work reported accuracy of 90% in the experiments where training set was 50% of the dataset. These results also substantiate the role played by newer incorporations in refining IDS.

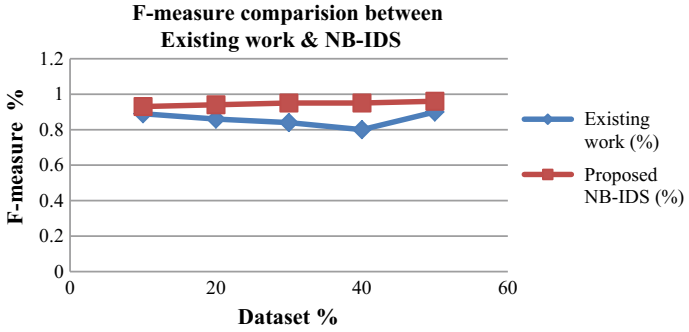

Fig. 4 Accuracy between existing and proposed NB-IDS

5 Conclusion

Intrusion detection systems strive to identify the anomalies by monitoring the incoming traffic but due to overgrowing and evolving attacks somehow it fails in certain conditions, also it suffers from problem of low training set size and update. This proposed approach combines rough set theory to identity intrusions more efficiently and effectively with fewer false alarms. The proposed integrates rough set theory in IDS to facilitate better identification of potential malicious contents. Experimental results clearly indicates NB-IDS with rough set reports better detection rate and lower false positive, subsequently outperforms the existing state of the art.

References

1. Deshpande, V.K.: Intrusion detection system using decision tree based attribute weighted AODE. Int. J. Adv. Res. Comput. Commun. Eng. **3**(12), 878–8743 (2014)
2. Rai, K., Devi, M.S., Guleria, P.: Decision tree based algorithm for intrusion detection. Int. J. Adv. Netw. Appl. **07**(04), 2828–2834 (2016)
3. Raman, M.R.G., Kannan, K., Pal, S.K., Shankar, V.: Rough set-hyper graph-based feature selection approach for intrusion detection systems. Def. Sci. J. **66**(6), 612–617 (2016)
4. Pawlak, Z.: Rough sets. Int. J. Comput. Inf. Sci. **11**(5), 341–356 (1982)
5. Li, T., Nguyen, H.S., Wang, G.Y.: Rough Sets and Knowledge Technology, pp. 369–378. Springer, Berlin, Heidelberg (2012)
6. Zhang, Q., Xie, Q., Wang, G.: A survey on rough set theory and its applications. CAAI Trans. Intell. Technol. **1**(4), 323–333 (2016)
7. Paul, P., Dey, U., Roy, P., Goswami, S.: Network security based on rough set: a review. J. Netw. Commun. Emerg. Technol. (JNCET) **5**(2), 165–169 (2015)
8. Thaseen, S., Kumar, A.: An analysis of supervised tree based classifiers for intrusion detection system. In: International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp. 294–299 (2013)
9. Ingre, B., Yadav, A., Soni, A.K.: Decision tree based intrusion detection system for NSL-KDD dataset. In: International Conference on Information and Communication Technology for Intelligent Systems ICTIS vol. 2, pp. 207–218 (2017)

10. Bordbar, S., Abdulah, M.B.T.: A feature selection based on rough set for improving intrusion detection system. Int. J. Eng. Sci. (IJES) **4**(7), 54–60 (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 the way ahead. In: Recent Trends in Computer Networks and Distributed Systems Security, pp. 474–484. Springer (2011)
12. Saurabh, P., Verma, B.: Cooperative negative selection algorithm. Int. J. Comput. Appl. (0975 – 8887) **95**(17), 27–32 (2014)
13. Saurabh, P., Verma, B., Sharma, S.: An immunity inspired anomaly detection system: a general framework. In: Proceedings of 7th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012), pp. 417–428. Springer (2012)
14. Saurabh, P., Verma, B.: An efficient proactive artificial immune system based anomaly detection and prevention system. Expert Syst. Appl. **60**, 311–320 (2016)