A NOVEL TECHNIQUE FOR INTRUSION DETECTION SYSTEM FOR NETWORK SECURITY USING HYBRID SVM-CART

Aastha Puri[1], Nidhi Sharma[2]

Research Scholar[1], Assistant Professor[2]

SDDIET Department of Computer Sc. Barwala Haryana, India

***Abstract---** Intrusion detection in the field of computer network is an important area of research from the past few years. Many approaches of classification have been proposed and their merits and demerits have been compared with the existing approaches. Machine learning approaches are used with the detection systems to accurately detect the attacks in real time. In the present approach a hybrid algorithm is proposed which is a combination of Support Vector Machine and Classification and Regression tree algorithm for classifying the attacks. The proposed technique is then applied on the KDDCUP 99 dataset. The proposed algorithm shows better results in terms of accuracy detection rate and false rate. In future other machine learning algorithms must be used and their results must be compared with the existing approaches.*

***Keywords:** Intrusion Detection System, SVM-CART, Classification*

INTRODUCTION

In information technology, a server is considered any instance of an application that can receive and serve the requests of other programs. Usually these applications are run on computers dedicated to acting solely as servers so that the heavy burden of fulfilling requests from other devices on the network does not overwhelm the computers. Running servers on dedicated computers is also a safety measure, helping to keep the server from being attacked [1]. The computers dedicated to acting as servers usually include faster CPUs, bigger hard drives, better RAM, and multiple power sources. These enhancements allow the server to handle the immense workload and also give it reliability in the event of unfortunate events.

As the internet evolves and computer networks become bigger and bigger, network security has become one of the most important factors for companies to consider. Big enterprises like Microsoft are designing and building software products that need to be protected against foreign attacks [2]. Anything from software, music and movies to books, games, etc. are stolen and copied because security is breached by malicious individuals. Today, most malicious users do not possess a high level of programming skills and instead make use of tools available on the Internet. There are several stages that an attacker has to pass through to successfully carry out an attack [3].

The Intrusion Detection System is capable of detecting the unwanted access to the system of organizations. To make it more efficient some heuristic approach should be taken. Researchers have been trying to come up with the better solution to this problem. Intrusion detection system is a system on which immense techniques can be combined and compared [4]. For example one can combine meta-heuristic based clustering with efficient feature selection technique. Hence, there is tremendous future scope in development of intrusion detection system, as one can consider the ups and downs of various techniques. Also, the proper blend of the useful techniques can result in an ideal or near to ideal intrusion detection system.Intrusion detection system has becoming a wide research area for the researchers to come up with a better algorithm to classify the intrusion on any system before blocking them. To achieve such real time, accurate and intelligent IDS, researchers are applying meta-heuristic techniques to IDS. Since there is a wide research going on in the field of meta-heuristic technique and IDS is always open to give a better result by applying such technique to it which are feasible to merge with IDS [5].

An IDS is software or may be referred as device which helps to monitor a system or network for an malicious activity or an kind of violations. It there may be any violation or any malicious activity detected then it is usually reported to the administrator. Any violation may also be reported to the security information and event management (SIEM) system.

This SIEM system uses alarm filtering approach to differentiate any violation from false alarm. This SIEM system combines its output from different sources. IDS system is wide range of spectrum which may vary from antivirus software to hierarchical systems that verify the traffic of whole backbone network [6]. The common categories of DS system are: network intrusion detection systems (NIDS) and host-based intrusion detection systems (HIDS). NIDS system is those which analyze incoming network traffic and on the other hand HIDS systems are those which checks significant operation systems only. Some IDS have the capability to react to detected intrusions. Systems with response capabilities are usually referred to as an intrusion prevention system.

RELAED WORK

Dubey, Shreya, and JigyasuDubey. [1] The proposed study is an investigation of IDS (intrusion detection system) and their design concept. For that purpose an intrusion detection system is developed using the analysis of KDD CUP 99's dataset. The main focus is given over classification and performance improvement of classifiers. The proposed IDS system utilizes the k-mean clustering algorithm, Bayesian classification algorithm and finally the back propagation neural network. The implementation of the desired system is performed using MATLAB IDE. In order to justify the performance outcomes, the proposed classification technique is compared with the traditional Bayesian classifier. According to the obtained performance the proposed classification technique provides optimum classification accuracy and error rate improvement over the traditional method but the performance is similar in terms of prediction time. On the other hand the memory consumption of the traditional method is better than the proposed method. Our proposed approach provides better accuracy then the conventional kmeans, k-nearest neighbor and naive-bayes. According to the obtained performance the system is adoptable and efficient. In near future the performance of the classification is improved more as reducing the steps of algorithm which is time consuming.

Tseng, Chin-Yang, PoornimaBalasubramanyam [2] propose a specification-based intrusion detection system that can detect attacks on the AODV routing protocol. In a specification-based intrusion detection approach, the correct behaviours of critical objects are manually abstracted and crafted as security specifications, and this is compared with the actual behaviour of the objects. Intrusions, which usually cause object to behaviour in an incorrect manner, can be detected without exact knowledge about them. This approach can, thus, address unknown attacks as well. The IDS presented in this paper is built on a distributed network monitor architecture that traces AODV request-reply flows. Network monitors audit every RREQ, RREP and RERR in order to build and update complete request-reply session trees and corresponding forwarding tables. Constraints on the request-reply flow are specified using finite state machines. It describes procedures for constructing and processing the session trees, and present examples of detecting attacks successfully. This research is the first effort to apply specification-based detection techniques to detect attacks in the routing within ad hoc networks. The work illustrate that our algorithm can effectively detect most of the serious AODV routing attacks effectively, and with low overhead.

Faisal, Mustafa Amir, Zeyar Aung[3] in this paper, proposed architecture for the comprehensive IDS in AMI, which is designed to be reliable, dynamic, and considering the real-time nature of traffic for each component in AMI. Then, it conducts a performance analysis experiment of the seven existing state-of-the-art data stream mining algorithms on a public IDS data set. Finally, it is elucidate the strengths and weaknesses of those algorithms and assess the suitability of each of them to serve as the IDSs for the three different components of AMI. This has been observed that some algorithms that use very minimal amount of computing resources and offer moderate level of accuracy can potentially be used for the smart meter IDS. On the other hand, the algorithms that require more computing resources and offer higher accuracy levels can be useful for the IDSs in data concentrators and AMI head ends.

Roesch, Martin [4] In this paper Snort was designed. This proposed design is used to fulfill the requirements of a prototypical lightweight network intrusion detection system. It has become a small, flexible, and highly capable system that is in use around the world on both large and small networks. It has attained its initial design goals and is a fully capable alternative to commercial intrusion detection systems in places where it is cost inefficient to install full featured commercial systems.

Debar, Herve, Monique Becker[5] in this paper, intrusion detection system has been proposed. The user model which is developed in this paper is the complement of a statistical model, because neural networks cannot adequately handle all the available data. The tight coupling between the neural net and the expert system is necessary to analyses the output of the net and propose explanations and a clear diagnosis to the security administrator.

Peddabachigari, Sandhya,[6] in this research, some new techniques for intrusion detection has been investigated and evaluated their performance based on the benchmark KDD Cup 99 Intrusion data. This work presented DT and SVM as intrusion detection models. Next, a hybrid DT–SVM model is designed and an ensemble approach with DT, SVM and DT–SVM models as base classifiers. Empirical results reveal that DT gives better or equal accuracy for Normal, Probe, U2R and R2L classes. The hybrid DT–SVM approach improves or delivers equal performance for all the classes when compared to a direct SVM approach. The Ensemble approach gave the best performance for Probe and R2L classes. The ensemble approach gave 100% accuracy for Probe class, and this suggests that if proper base classifiers are chosen 100% accuracy might be possible for other classes too. Finally, proposes a hierarchical intelligent IDS model to make optimum use of the best performances delivered by the individual base classifiers and the ensemble approach.

Shah, Bhavin, and Bhushan H. Trivedi.[7] This paper discussed reasons and solutions for these challenges and verified these solutions on computer network having 1400 systems by performing various experiments using Jade Agent Platform. These experiments clearly show that our solutions reduce response time and agent size to a great extent.

Rosenberg, Ishaiet al**. [8]** proposed Intrusion detection system based on the system calls.IDS has becoming a wide research area for the researchers to come up with a better algorithm to classify the intrusion on any system before blocking them. Moreover in this paper, researchers are applying meta-heuristic techniques to IDS in order to achieve such real time, accurate and intelligent IDS. Since there is a wide research going on in the field of meta-heuristic technique and IDS is always open to give a better result by applying such technique to it which are feasible to merge with IDS.

Nápoles, Gonzalo, IselGrau[9] introduced a novel IDS based on Rough Cognitive Networks, a recently proposed granular neural network for pattern classification. Without loss of generality, we can define RCN as a Sigmoid Fuzzy Cognitive Map where input neurons represent information granules whereas output concepts denote decision classes. It should be remarked that the granulation of information is achieved by using Rough Sets, since it allows handling uncertainty arising from inconsistency. Furthermore, with the goal of increasing the reliability of the RCN-based inference process, and discussed a supervised learning methodology for automatically computing accurate similarity relations by estimating the proper parameter vector.

PROPOSED METHODOLOGY

The proposed work aims to develop a hybrid algorithm of Linear Discriminant Analysis based Support Vector Machine-Classification and Regression Tree. The SVM classifier classifies different data into special clusters relied on the values of support vector and it uses a support vector along the main component. The Algorithm will be hybrid with CART algorithm that is depended on the regression tree concept. CART is classification method which uses historical data to construct decision trees. Classification tree or regression tree may be constructed which is depending on the information available about dataset. These regression trees may be utilized for the classification of the new observation. Classes in learning sample can be given by user or evaluated in accordance with some exogenous rule. The LDA is utilized for extracting feature from the huge number of features that will decrease the cost of computation.

Steps

1. Data is obtained from KDD.
2. Linear Discriminant Analysis is applied for reduction of dimensionality of the data and selects the best features out of it.
3. Classification algorithm will be applied on the data.
4. SVM is given more weightage for nearby points to the line and CART is given more weightage for distant points.
5. Activities will be monitored and parameters will be calculated Figure 1 shows the flow diagram of the proposed methodology.

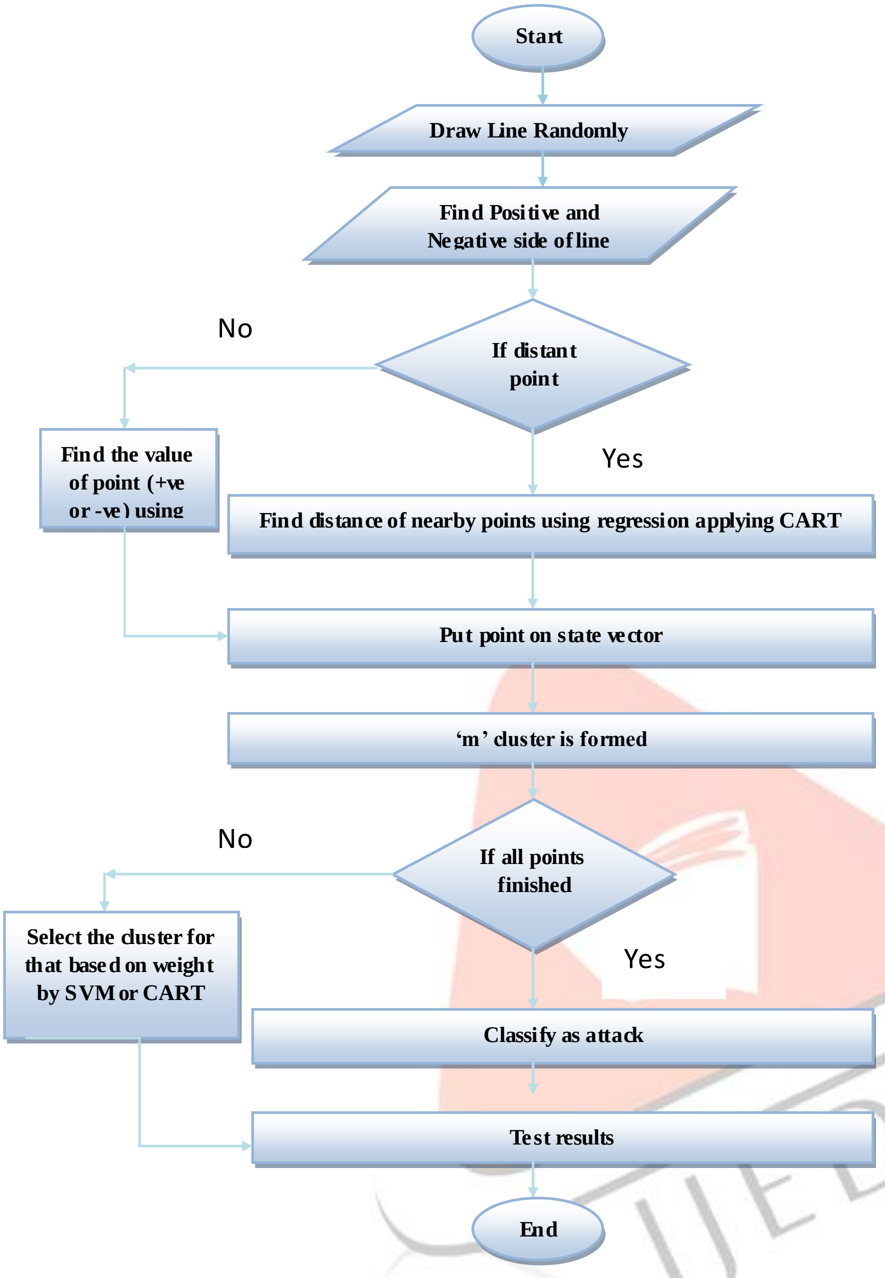
Start
Draw Line Randomly
Find Positive and Negative side of line
If distant point
No
Yes
Find the value of point (+ve or -ve) using
Find distance of nearby points using regression applying CART
Put point on state vector
'm' cluster is formed
If all points finished
No
Yes
Select the cluster for that based on weight by SVM or CART
Classify as attack
Test results
End

Fig 1: Flow Diagram of Proposed Methodology

RESULTS AND DISCUSSIONS

The proposed algorithm of SVM-CART has been implemented along with KNN for comparing the two. For assessment of attacks over the network for both algorithms have implemented in python and have been tested on KDD CUP 99 dataset. The result after performing test on KDD dataset by both the algorithm is shown in figures below. Figure 1 shows the Accuracy graph of the proposed algorithm of SVM-CART. It can be concluded that the overall accuracy of the algorithm is above 95 percent except the U2R which depends on the available training set from KDD dataset. The algorithm is been tested on the 10 percent training set of actual KDD dataset. The reason to take 10% of the training set is execution time. If the proposed algorithm will be tested taking all dataset as training set result will be much better.

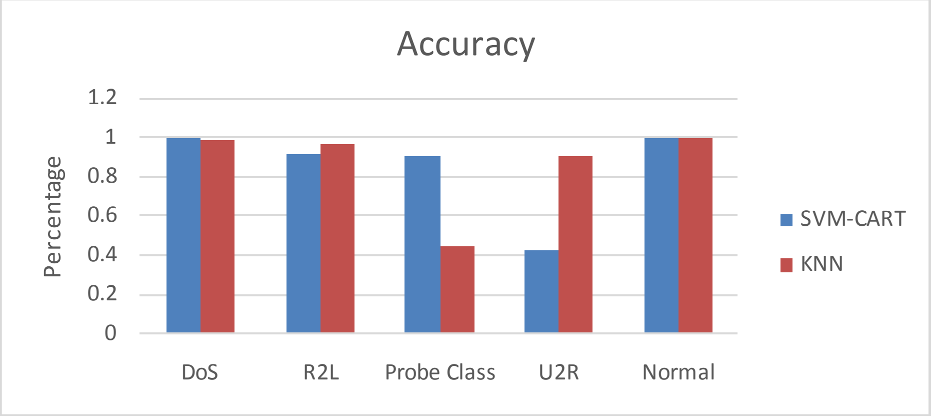

Fig1: Accuracy using a) SVM-CART and b) KNN

Figure1 (a) represents the graph of accuracy of matched labels of classified classes. Figure 1 (b) shows the accuracy of using KNN. Here X-axis represents different class of attacks and Y-axis represents the percentage. On X-axis Normal class, Denial of Services (DoS), User-to-Remote (U2R), User to root (R2L) and Probing class has been defined. From above graph it has been clear that the SVM algorithm is quite capable of clustering the points as accurate as possible as and better than KNN as we can see it is able to classify DOS more accurately. As we can see that most of the classes matched is more than 95% accurate.

Table 1 shows the values of Accuracy for SVM-CART and KNN.

Class of Attacks	SVM-CART	KNN
DoS	1	0.99
R2L	0.92	0.965
Probe Class	0.91	0.45
U2R	0.43	0.91
Normal	1	1

Table 1 Accuracy

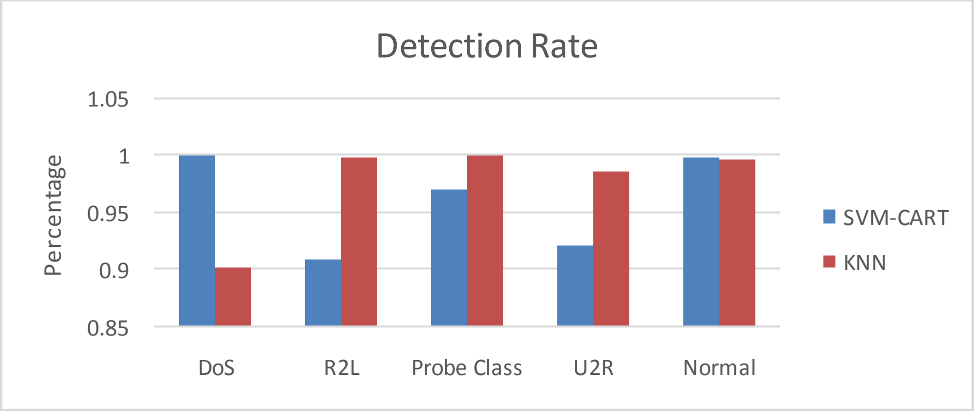

Fig 2: Detection rate with a) SVM-CART and b) KNN

Figure 2 (a) shows the detection rate of SVM-CART and Figure 2 (b) shows the detection rate of KNN. Table 2 shows the detection rate values of SVM_CART and KNN. The over-all Detection Rate of the proposed algorithm for the given dataset is coming out to be more than 0.95.it shows the detection rate of individual classes of attacks with the proposed algorithm for the given dataset.

Class of Attacks	SVM-CART	KNN
DoS	1	0.902
R2L	0.91	0.998
Probe Class	0.97	1
U2R	0.922	0.986
Normal	0.998	0.996

Table 2 Detection

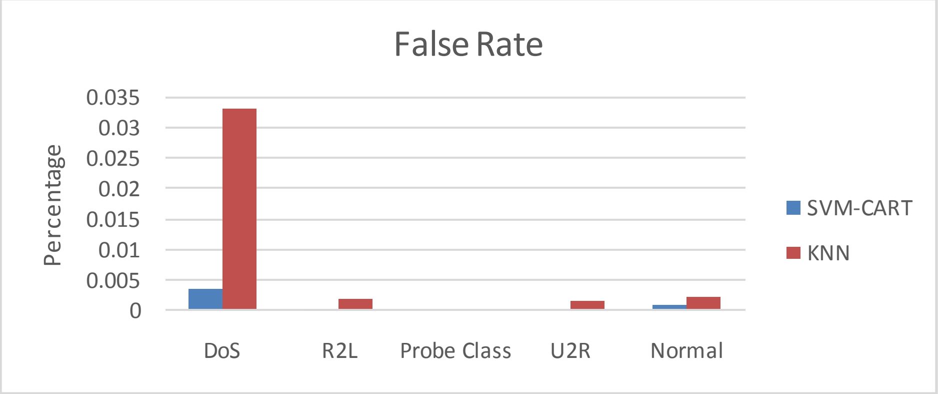

Fig3: False rate with a) SVM-CART and b) KNN

Figure 3 (a)shows the false rate of the different classes of attacks with SVM-CART and Figure 3(b) shows the false rate of the different classes of attacks with KNN from the mentioned graph is been clear that the value of false rate is coming out to be less than 0.035 which is 3.5% in case of SVM-CART of the original dataset and it is known that lesser is the false rate better is the algorithm. Table 3 shows the false rate for different class of attacks.

Class of Attacks	SVM-CART	KNN
DoS	0.0036	0.033
R2L	0.0003	0.002
Probe Class	0.0002	0
U2R	0.00005	0.0015
Normal	0.001	0.0024

Table 5.3 False Rate

CONCLUSION

The proposed technique of SVM-CART (Support Vector Machine – Classification and Regression Testing) has been applied to the set of KDDCUP 99 dataset along with the KNN for the comparison purpose. Intrusion detection is the process of detecting and classifying the attacks performed over the network of system. By achieving this assessment to more than 90 percent we are protecting or making a network of system more secure or resistant to attacks. KDDCUP dataset is the set of all attacks performed using high end infrastructure which is been used to train our proposed algorithm and then tested to check the result of the proposed algorithm. It has been found that the proposed technique has provided a promising result even on taking 10% of the actual dataset as training set to the algorithm and then tested. The proposed algorithm can be used to achieve the realistic result by appending it to the real network analytic tool used to capture packet transferred over the network such as Wire-Shark, Snort etc.

REFERENCES

[1] Dubey, Shreya, and JigyasuDubey. "KBB: A hybrid method for intrusion detection." In Computer, Communication and Control (IC4), 2015 International Conference on, pp. 1-6. IEEE, 2015.

[2] Tseng, Chin-Yang, PoornimaBalasubramanyam, Calvin Ko, RattaponLimprasittiporn, Jeff Rowe, and Karl Levitt. "A specification-based intrusion detection system for AODV." In Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks, pp. 125-134. ACM, 2003.

[3] Faisal, Mustafa Amir, Zeyar Aung, John R. Williams, and Abel Sanchez. "Data-stream-based intrusion detection system for advanced metering infrastructure in smart grid: A feasibility study." Systems Journal, IEEE 9, no. 1 (2015): 31-44.

[4] Roesch, Martin. "Snort: Lightweight Intrusion Detection for Networks." InLISA, vol. 99, no. 1, pp. 229-238. 2014.

[5] Debar, Herve, Monique Becker, and Didier Siboni. "A neural network component for an intrusion detection system." In Research in Security and Privacy, 1992. Proceedings., 1992 IEEE Computer Society Symposium on, pp. 240-250. IEEE, 1992.

[6] Peddabachigari, Sandhya, Ajith Abraham, CrinaGrosan, and Johnson Thomas. "Modeling intrusion detection system using hybrid intelligent systems." Journal of network and computer applications 30, no. 1 (2007): 114-132.

[7] Shah, Bhavin, and Bhushan H. Trivedi. "Improving Performance of Mobile Agent Based Intrusion Detection System." In Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference on, pp. 425-430. IEEE, 2015.

[8] Rosenberg, Ishai, and Ehud Gudes. "Evading System-Calls Based Intrusion Detection Systems." In *International Conference on Network and System Security*, pp. 200-216. Springer International Publishing, 2016.

[9] Nápoles, Gonzalo, IselGrau, Rafael Falcon, Rafael Bello, and Koen Vanhoof. "A Granular Intrusion Detection System Using Rough Cognitive Networks." In Recent Advances in Computational Intelligence in Defense and Security, pp. 169-191. Springer International Publishing, 2016.