IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Network Intrusion Detection By SVM & ANN With Feature Selection

[1]B.VenkataRamana,[2]K Chandra Mouli,[3]Aileni Eenaja

[1,2,3]Assistant Professor

[1 ,2,3]Computer Science and Engineering

[1,2,3]Holy Mary Institute of Technology & Science, Hyderabad,India

Abstract: A Supervised machine learning system is to classify network traffic whether it is malicious or benign. To find the best model considering detection success rate, combination of supervised learning algorithm and feature selection method have been used. Through this study, it is found that Artificial Neural Network (ANN) based machine learning with wrapper feature selection outperform support vector machine (SVM) technique while classifying network traffic. To evaluate the performance, NSL-KDD dataset is used to classify network traffic using SVM and ANN supervised machine learning techniques. Comparative study shows that the proposed model is efficient than other existing models with respect to intrusion detection success rate. Specially at present technology playing major role so we need to move forward with our upgrading mindset

Index Terms **- SVM, ANN, NSL-KDD.**

I. Introduction

Intrusion detection is the first step to prevent security attack. Hence the security solutions such as Firewall, Intrusion Detection System (IDS), Unified Threat Modeling (UTM) and Intrusion Prevention System (IPS) are getting much attention in studies. IDS detects attacks from a variety of systems and network sources by collecting information and then analyzes the information for possible security breaches . The network based IDS analyzes the data packets that travel over a network and this analysis are carried out in two ways. Till today anomaly based detection is far behind than the detection that works based on signature and hence anomaly based detection still remains a major area for research . The challenges with anomaly based intrusion detection are that it needs to deal with novel attack for which there is no prior knowledge to identify the anomaly. Hence the system somehow needs to have the intelligence to segregate which traffic is harmless and which one is malicious or anomalous and for that machine learning techniques are being explored by the researchers over the last few years . IDS however is not an answer to all security related problems. For example, IDS cannot compensate weak identification and authentication mechanisms or if there is a weakness in the network protocols. Studying the field of intrusion detection first started in 1980 and the first such model was published in 1987 . For the last few decades, though huge commercial investments and substantial research were done, intrusion detection technology is still immature and hence not effective . While network IDS that works based on signature have seen commercial success and widespread adoption by the technology based organization throughout the globe, anomaly based network IDS have not gained success in the same scale. Due to that reason in the field of IDS, currently anomaly based detection is a major focus area of research and development . And before going to any wide scale deployment of anomaly based intrusion detection system, key issues remain to be solved. But the literature today is limited when it comes to compare on how intrusion detection performs when using supervised machine learning techniques. To protect target systems and networks against malicious activities anomaly-based network IDS is a valuable technology. Despite the variety of anomaly-based network intrusion detection techniques described in the literature in recent years , anomaly detection functionalities enabled security tools are just beginning to appear, and some important problems remain to be solved. Several anomaly based techniques have been proposed including Linear Regression, Support Vector Machines (SVM), Genetic Algorithm, Gaussian mixture model, knearest 3704abelling algorithm, Naive Bayes classifier, Decision Tree. Among them the most widely used learning algorithm is SVM as it has already established itself on different types of problem. One major issue on anomaly based detection is though all these proposed techniques can detect novel attacks but they all suffer a high false alarm rate in general. The cause behind is the complexity of generating profiles of practical normal behaviour by learning from the training data sets. Today Artificial Neural Network (ANN) are often trained by the back propagation algorithm, which had been around since 1970 as the reverse mode of automatic differentiation.The major challenges in evaluating performance of network IDS is the unavailability of a comprehensive network based data set. Most of the proposed anomaly based techniques found in the literature were evaluated using KDD CUP 99 dataset. In this paper we used SVM and ANN –two machine learning techniques, on NSLKDD which is a popular benchmark dataset for network intrusion.

Algorithms Used:

Two supervised machine learning algorithms such as SVM (Support Vector Machine) and ANN (Artificial Neural Networks). Machine learning algorithms will be used to detect whether request data contains normal or attack (anomaly) signatures. Now-a-days all services are available on internet and malicious users can attack client or server machines through this internet and to avoid such attack request IDS (Network Intrusion Detection System) will be used, IDS will monitor request data and then check if its contains normal or attack signatures, if contains attack signatures then request will be dropped. IDS will be trained with all possible attacks signatures with machine learning algorithms and then generate train model, whenever new request signatures arrived then this model applied on new request to determine whether it contains normal or attack signatures. In this paper we are evaluating performance of two machine learning algorithms such as SVM and ANN and through experiment we conclude that ANN outperform existing SVM in terms of accuracy. To avoid all attacks IDS systems has developed which process each incoming request to detect such attacks and if request is coming from genuine users then only it will forward to server for processing, if request contains attack signatures then IDS will drop that request and log such request data into dataset for future detection purpose. To detect such attacks IDS will be prior train with all possible attacks signatures coming from malicious user's request and then generate a training model. Upon receiving new request IDS will apply that request on that train model to predict it class whether request belongs to normal class or attack class. To train such models and prediction various data mining classification or prediction algorithms will be used.

Evaluating performance of SVM and ANN.

Algorithms has applied Correlation Based and Chi-Square Based feature selection algorithms to reduce dataset size, this feature selection algorithms removed irrelevant data from dataset and then used model with important features, due to this features selection algorithms dataset size will reduce and accuracy of prediction will increase.

II. Literature Survey

Recently, Internet has become a part and parcel of daily life. The current internet-based information processing systems are prone to different kinds of threats which lead to various types of damages resulting in significant losses. Therefore, the importance of information security is evolving quickly. The most basic goal of information security is to develop defensive information systems which are secure from unauthorized access, use, disclosure, disruption, modification, or destruction. Moreover, information security minimizes the risks related to the three main security goals namely confidentiality, integrity, and availability.Various systems have been designed in the past to identify and block the Internet-based attacks. The most important systems among them are intrusion detection systems (IDS) since they resist external attacks effectively. Moreover, IDSs provide a wall of defence which overcomes the attack of computer systems on the Internet. IDS could be used to detect different types of attacks on network communications and computer system usage where the traditional firewall cannot perform well. Intrusion detection is based on an assumption that the behaviour of intruders differ from a legal user. Generally, IDSs are broadly classified into two categories namely anomaly and misuse detection systems based on their detection approaches. Anomaly intrusion detection determines whether deviation from the established normal usage patterns can be flagged as intrusions. On the other hand, misuse detection systems detect the violations of permissions effectively. Intrusion detection systems can be built by using intelligent agents and classification techniques. Most IDSs work in two phases namely preprocessing phase and intrusion detection phase. The intrusions identified by the IDSs can be prevented effectively by developing an intrusion prevention system. This paper mainly provides a survey on intelligent techniques proposed for developing IDSs. In addition, it explains about new IDS which has been developed using two proposed algorithms namely intelligent rule-based attribute selection algorithm and intelligent rule-based enhanced multiclass support vector machine (IREMSVM).

Intelligent IDSs are the ones considered to be intelligent computer programs situated in either a host or a network which analyzes the environment and acts flexibly to achieve higher detection accuracy . These programs compute the actions to be performed on the environment both by learning the environment and by firing rules of inference. Intelligent IDSs are capable of decision making and constraint checking. In most intelligent systems, either rules are fired or agents are used for decision making. Moreover, a set of static agents or a set of mobile and static agents have been used to achieve a single goal. Intelligent intrusion detection systems have been developed by proposing intelligent techniques for preprocessing and effective classification. Such IDSs have provided better detection rate in comparison with the other approaches.

"A macro-social exploratory analysis of the rate of interstate cyber-victimization" :

This study examines whether macro-level opportunity indicators affect cyber-theft victimization. Based on the arguments from criminal opportunity theory, exposure to risk is measured by state-level patterns of internet access (where users access the internet). Other structural characteristics of states were measured to determine if variation in social structure impacted cyber-victimization across states. The current study found that structural conditions such as unemployment and non-urban population are associated with where users access the internet. Also, this study found that the proportion of users who access the internet only at home was positively associated with state-level counts of cyber-theft victimization. The theoretical implications of these findings are discussed.

"Incremental anomaly-based intrusion detection system using limited labeled data":

With the proliferation of the internet and increased global access to online media, cybercrime is also occurring at an increasing rate. Currently, both personal users and companies are vulnerable to cybercrime. A number of tools including firewalls and Intrusion Detection Systems (IDS) can be used as defence mechanisms. A firewall acts as a checkpoint which allows packets to pass through according to predetermined conditions. In extreme cases, it may even disconnect all network traffic. An IDS, on the other hand, automates the monitoring process in computer networks. The streaming nature of data in computer networks poses a significant challenge in building IDS. In this paper, a method is proposed to overcome this problem by performing online classification on datasets. In doing so, an incremental naive Bayesian classifier is employed. Furthermore, active learning enables solving the problem using a small set of labeled data points which are often very expensive to acquire. The proposed method includes two groups of actions i.e. offline and online. The former involves data preprocessing while the latter introduces the NADAL online method. The proposed method is compared to the incremental naive Bayesian classifier using the NSL-KDD standard dataset. There are three advantages with the proposed method: (1) overcoming the streaming data challenge; (2) reducing the high cost associated with instance labeling; and (3)

improved accuracy and Kappa compared to the incremental naive Bayesian approach. Thus, the method is well-suited to IDS applications.

"Modelling and implementation approach to evaluate the intrusion detection system"

Intrusions detection systems (IDSs) are systems that try to detect attacks as they occur or when they were over. Research in this area had two objectives: first, reducing the impact of attacks; and secondly the evaluation of the system IDS. Indeed, in one hand the IDSs collect network traffic information from some sources present in the network or the computer system and then use these data to enhance the systems safety. In the other hand, the evaluation of IDS is a critical task. In fact, it's important to note the difference between evaluating the effectiveness of an entire system and evaluating the characteristics of the system components. In this paper, we present an approach for IDS evaluating based on measuring the performance of its components. First of all, in order to implement the IDS SNORT components safely we have proposed a hardware platform based on embedded systems. Then we have tested it by using a generator of traffics and attacks based on Linux KALI (Backtrack) and Metasploite 3 Framework. The obtained results show that the IDS performance is closely related to the characteristics of these components.

III. System Architecture

System Architecture or Systems Architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. It can consist of system components and the sub-systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture; collectively these are called architecture description language The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

Feature Selection:

Feature selection is an important part in machine learning to reduce data dimensionality and extensive research carried out for a reliable feature selection method. For feature selection filter method and wrapper method have been used. In filter method, features are selected on the basis of their scores in various statistical tests that measure the relevance of features by their correlation with dependent variable or outcome variable. Wrapper method finds a subset of features by measuring the usefulness of a subset of feature with the dependent variable. Hence filter methods are independent of any machine learning algorithm whereas in wrapper method the best feature subset selected depends on the machine learning algorithm used to train the model. In wrapper method a subset evaluator uses all possible subsets and then uses a classification algorithm to convince classifiers from the features in each subset. The classifier consider the subset of feature with which the classification algorithm performs the best. To find the subset, the evaluator uses different search techniques like depth first search, random search, breadth first search or hybrid search. The filter method uses an attribute evaluator along with a ranker to rank all the features in the dataset. Here one feature is omitted at a time that has lower ranks and then sees the predictive accuracy of the classification algorithm. Weights or rank put by the ranker algorithms are different than those by the classification algorithm. Wrapper method is useful for machine learning test whereas filter method is suitable for data mining test because data mining has thousands of millions of features.

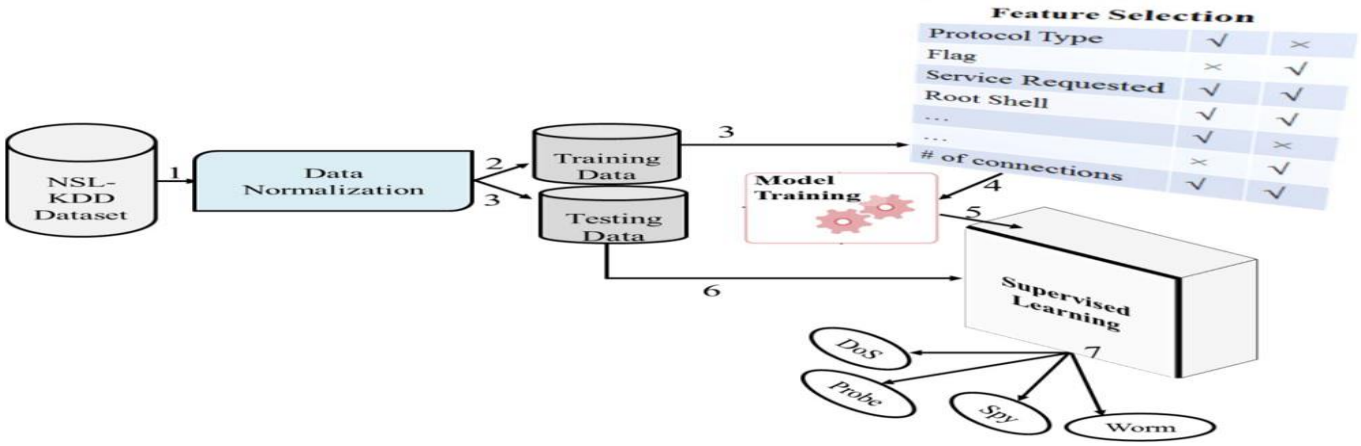

Fig : Architecture of NSL-KDD for Feature Selection

Classification:

With the features found in feature selection part, total four models are built in Weka software suite using the training dataset. Classification using supervised machine learning first requires training the model using training dataset. We used 20% of NSL-KDD dataset as training data. That have 25,191 labelled data instances. To training the model we used SVM and ANN learning algorithm for each type of feature selection method. Hence we build four learning models, two model using SVM and another 2 using ANN. Among the 2 model built for each learning algorithm, one is built using 17 features and another one is built using 35 features found in the feature selection part. Next these four trained models were evaluated using 22,542 instances of testing data picked from the NSL-KDD testing dataset. The findings are summarized in Table below. We listed our results with recently published results in the literature. While comparing the performance of the proposed model with the others works, we picked works having hypothesis of comparable aspects related to learning algorithm and benchmarking datasets. But there are other aspects like attribute reduction, number of instances, the

number layers and learning rates used. The detection success rate of the proposed model is also compared with other existing models in Table below.

Learning Type	Our Model Accuracy	Existing Model	Existing Model
SVM	82.34%	92.84% [16]	69.52% [17]
ANN	94.02%	81.2% [18]	77.23% [19]

Table: Performance Comparison with Existing Models

IV. Existing System

There is no staff available in unmanned restaurants, it is difficult for the restaurant management to estimate how the concept and the food is experienced by the customers. Existing rating systems, such as Google and Trip Advisor, only partially solve this problem, as they only cover a part of the customer's opinions. These rating systems are only used by a subset of the customers who rate the restaurant on independent rating platforms on their own initiative. This applies mainly to customers who experience their visit as very positive or negative. Issues are been stated in the existing literature survival like additional training time, accurate identification of low common attacks and attacks classification. In order to solve the issue of additional training time, it is must to develop a new high-speed algorithm for intrusion detection system and its results will be tested with existing techniques. In contrast to the existing approaches that performed some kind of inefficiency in intrusion detection, the main aim of our research work is to propose a new high speed algorithm for reducing training time. Obtained results are also to be discussed along with the existing methods.

V. Proposed System

Here we solve the above problems raised in real time, customers must be motivated to give a rating. We introduces an approach for a restaurant rating system that asks every customer for a rating after their visit to increase the number of ratings as much as possible. This system can be used unmanned restaurants; the scoring system is based on facial expression detection using pertained convolution neural network (CNN) models. It allows the customer to rate the food by taking or capturing a picture of his face that reflects the corresponding feelings. Compared to text-based rating system, there is much less information and no individual experience reports collected. This simple fast and playful rating system should give a wider range of opinions about the experiences of the customers with the restaurant concept. we will be able to solve additional training time issues, and develop a new high-speed algorithm for intrusion detection system and its result will be tested with existing techniques to get accurate result. In contrast to the older approaches which are performed have some kind of inefficiency in intrusion detection, so it can be solved by applying ANN. The main aim is to give more accuracy by applying high speed algorithm and reduce the training time. Thus we can get perfect accuracy and we have to compare to see which gave better accuracy.

VI. Conclusion

We have presented different machine learning models using different machine learning algorithms and different feature selection methods to find a best model. The analysis of the result shows that the model built using ANN and wrapper feature selection outperformed all other models in classifying network traffic correctly with detection rate more compare to existing models. Our belief is that these findings will contribute to research further in the domain of building a detection system that can detect known attacks as well as novel attacks. The intrusion detection system exist today can only detect known attacks. Detecting new attacks or zero day attack still remains a research topic due to the high false positive rate of the existing systems.

References

[1] H. Song, M. J. Lynch, and J. K. Cochran, "A macro-social exploratory analysis of the rate of interstate cyber-victimization," American Journal of Criminal Justice, vol. 41, no. 3, pp. 583–601, 2016.

[2] P. Alaei and F. Noorbehbahani, "Incremental anomaly-based intrusion detection system using limited labelled data," in Web Research (ICWR), 2017 3th International Conference on, 2017, pp. 178–184.

[3] M. Saber, S. Chadli, M. Emharraf, and I. El Farissi, "Modeling and implementation approach to evaluate the intrusion detection system," in International Conference on Networked Systems, 2015, pp. 513–517.

[4] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward credible evaluation of anomaly-based intrusion-detection methods," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 5, pp. 516–524, 2010.

[5] A. S. Ashoor and S. Gore, "Importance of intrusion detection system (IDS)," International Journal of Scientific and Engineering Research, vol. 2, no. 1, pp. 1–4, 2011.

[6] M. Zamani and M. Movahedi, "Machine learning techniques for intrusion detection," arXiv preprint arXiv:1312.2177, 2013.

[7] N. Chakraborty, "Intrusion detection system and intrusion prevention system: A comparative study," International Journal of Computing and Business Research (IJCBR) ISSN (Online), pp. 2229–6166, 2013.

[8] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," computers & security, vol. 28, no. 1–2, pp. 18–28, 2009.

[9] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," Procedia Computer Science, vol. 89, pp. 117–123, 2016.

[10] J. Zheng, F. Shen, H. Fan, and J. Zhao, "An online incremental learning support vector machine for large-scale data," Neural Computing and Applications, vol. 22, no. 5, pp. 1023–1035, 2013.

[11] F. Gharibian and A. A. Ghorbani, "Comparative study of supervised machine learning technique