International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 26th Apr 2022. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-11&issue= Spl Issue 04

DOI: 10.48047/IJIEMR/V11/SPL ISSUE 04/10

Title **DETECTION OF INTRUSION ATTACK USING**

CUSTOMISED ANN

Volume 11, SPL ISSUE 04, Pages: 94-103

Paper Authors
V. Sowjanya, M. Ajith Bala Chandra Reddy, CH. Prabhu Nikhil Raj, P. Eswar Kamal

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

DETECTION OF INTRUSION ATTACK USING CUSTOMISED ANN

[1]V. Sowjanya, [2]M. Ajith Bala Chandra Reddy, [3]CH. Prabhu Nikhil Raj, [4]P. Eswar Kamal
[1]Associate Professor CSE, PSCMR College of Engineering & Technology, Vijayawada,
Andhra Pradesh
[2,3,4] Student, CSE, PSCMRCET, Vijayawada, Andhra Pradesh
sowjanyav@pscmr.ac.in, ajithreddy610@gmail.com, challanikhilraj@gmail.com, eswarkamal.p@gmail.com .

Abstract

Malicious assaults on client or server machines via the internet are more common than ever, and there is no way to prevent them. To address this, we suggest a method in which an IDS is requested (Network Intrusion Detection System). IDS will keep an eye on the provided data and determine whether it contains normal or malicious signatures. The request will be dropped if it is discovered to contain attack signatures. When new request signatures arrive, the IDS will be taught with all of the options for attacking signatures with machine learning techniques and then produce a training model, which will be used to future requests to determine whether they contain normal or attack signatures. We used machine learning methods such as SVM and ANN, as well as experimentation, to come up with a solution in terms of accuracy, the ANN outperforms the conventional SVM.

To avoid all attacks, IDS structures have evolved in such a way that every incoming request is analysed to locate such assaults, and if the request is coming from genuine users, it will only be sent to the server for processing; if the request contains assault signatures, the IDS will drop the request and log such request statistics into a dataset for future detection. To find such attacks, IDS might be taught with all possible assault signatures resulting from a malicious user's request first, then generate a schooling version. When IDS receives a new request, it will be practised on that teach version to determine whether the request belongs to ordinary elegance or assault elegance. Many statistics mining classes or prediction algorithms could be utilised to teach such fashions and predictions.

Keywords: Network Security, attacks, hackers, Cloud-environment security, zero-trust model (ZTM), Trend Micro internet security.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

1. INTRODUCTION

Because of the widespread use of the internet, there will be more access to online content, and cybercrime will be on the rise. The first step in preventing a safety assault is to detect intrusion. As a result, security solutions such as Firewall, Intrusion Detection System (IDS), Unified Threat Modeling (UTM), and Intrusion Prevention System (IPS) are gaining a lot of attention in research. IDS identifies attacks on a variety of structures and community reassets by collecting statistics and analysing them for possible security breaches. The community-based IDS analyses the data packets that travel over a network and performs this analysis in a variety of methods. Until date, anomaly-based detection has been lagging behind detection that is mostly based on data. Detection, on the other hand, remains a key area of study. The difficulties with anomaly-based completely intrusion detection are that it must deal with novel attacks for which there is no prior expertise to detect the abnormality. As a result, the gadget, in some way or another, wishes to have the intelligence to distinguish between which visitors are harmless and which are malevolent or abnormal, and researchers have been exploring system mastery tactics for the past few years. IDS, however, isn't always the case a All safety-related issues will be solved. DS, for example, cannot compensate for vulnerable identification and authentication mechanisms or if the community protocols have a flaw.

In 1980, researchers began researching the topic of intrusion detection, and the first model was published in 1987. Despite massive commercial investments and extensive research over the last few decades, intrusion detection technology remains immature and thus ineffective. While signature-based network IDS have found commercial success and widespread acceptance by technology-based organisations around the world, Anomaly-based network IDS haven't had the same level of success. As a result, anomaly-based detection is currently a key focal area of research and development in the field of IDS. And, before deploying an anomaly-based intrusion detection system on a large scale, critical challenges must be resolved. However, when it comes to comparing how intrusion detection performs when utilising supervised machine learning approaches, the literature is currently restricted. Anomaly-based network IDS is a valuable solution for protecting target computers and networks from malicious actions.

Despite the popularity of anomaly-based completely community intrusion detection systems described in the literature in recent years, anomaly detection capabilities provided by security hardware are only now beginning to materialise, and a few key issues remain to be resolved Linear regression, support vector machines (SVM), Genetic Algorithm, Gaussian aggregation model, okay closest neighbour set of rules, Naive Bayes classifier, and Decision tree are some of the completely anomaly-based algorithms that have been proposed. SVM is

the most widely utilised studying set of rules because it has already established itself on specific types of problems. Although all proposed algorithms can detect novel attacks, they all have a high false alarm rate. The reason for this is the difficulty of creating profiles of sensible regular behaviour with the use of learning from schooling data sets. Today, Artificial Neural Networks (ANN) are routinely trained utilising the returned propagation set of rules, which have been used since 1970 as the alternative form of computerised differentiation.

2. RELATED WORK

[1] Mubarak Albarka Umar et.al. The fundamental problem of machine learning-based IDS (Intrusion Detection System) with this method is that it takes a long time to compute. This is mostly due to superfluous, irrelevant, and incomplete data in the IDS datasets. The suggested system combines an IDS modelling technique with a feature selection algorithm (FS). Moustafa and Saly employed the UNSW-NB15 dataset, which has a 97.95 percent accuracy rating in the system when compared to other approaches. At first, there was a difficulty with utilising ML as a single classifier in IDS, however this model isn't good enough to develop a powerful IDS. To improve efficiency, the hybrid IDS modelling algorithm is utilised to analyse a combination of both feature selection techniques in the classification process. Because most datasets contain categorical features and the decision tree can handle both categorical and numeric features, this method suggests using a wrapper-based feature selection strategy with a decision tree algorithm as the feature evaluator. The major issue observed in all of the systems developed in this work is the high number of false alert rates, which will be rather tedious to work with if any of the models is to be deployed. A strong IDS should have a very low false alert rate, so more work can be done in the future focusing specifically on reducing the high false alert rate observed.

[2] Mohammed A. Ambusaidi et.al. The IDS (Intrusion Detection System) performs a main position in gazing and checking the sports of a device to locate the safety threats. The reason of this observe is to test and pick the extra discriminate the capabilities create the computationally powerful and green plan for an IDS. In this version, a hybrid function choice set of rules in aggregate of each clear out out and wrapper choice system is designed withinside the version. Least Square Support Vector Machine (LSSVM) is used in the system to manual and hold the set of capabilities. The important purpose for the device is to get most accuracy and minimal fake rate. Generally, IDS offers with big quantities of community statistics and unique forms of visitors patterns. The proposed device is the aggregate of clear out out function rating; and Improved Forward Floating Selection (IFFS) wrapper primarily based totally the use of LSSVM and classifying accuracy. The clear out out technique is used to lessen the computational fee and the wrapper seek is used to get rid of the redundancy and inappropriate capabilities from the preliminary function set. The framework of proposed hybrid function choice technique carries layers, top section wherein not

unusual place statistics is for removing and rating and decrease section which includes the most appropriate subset, and give you most accuracy on education dataset. In this version making use of the KDD ninety nine dataset to assess the overall working of recommend identify device. It selects randomly 15,246 data from the 2 unique training because the education statistics and the last 478,775 (494,021 - 15,246) samples are used for assessment purposes. The proposed detection version receives experimental effects as DR is ninety nine.47, and FP is 0.521. In order to destiny observe each proposed and new dataset may be used.

[3] Koushal Kumar et.al. The IDS (Intrusion Detection System) plays a first-rate function in watching and checking the sports activities of a tool to find the protection threats. The motive of this examine is to check and select out the more discriminate the competencies create the computationally effective and inexperienced plan for an IDS. In this model, a hybrid feature preference set of regulations in mixture of every clean out out and wrapper preference gadget is designed withinside the model. Least Square Support Vector Machine (LSSVM) is used in the selection gadget to control and preserve the set of competencies. The critical reason for the tool is to get maximum accuracy and minimum faux rate. Generally, IDS gives with large portions of network facts and specific types of traffic patterns. The proposed tool is the mixture of clean out out feature score; and Improved Forward Floating Selection (IFFS) wrapper based totally definitely the usage of LSSVM and classifying accuracy. Cleanup approach is used to reduce computational speed and wrapperResearch is used to defer redundancy and irrelevant skills from the initial feature set. The framework of proposed hybrid feature preference approach carries layers, pinnacle segment in which now no longer unusualplace facts is for getting rid of and score and reduce segment which incorporates the maximum suitable subset, and provide you with maximum accuracy on schooling dataset. In this model utilizing the KDD 99 dataset to evaluate the general overall working of propose identification tool. It selects randomly 15,246 statistics from the two specific schooling due to the fact the schooling facts and the ultimate 478,775 (494,021 - 15,246) samples are used for assesss causes. The proposed detection model gets experimental consequences as DR is 99.47, and FP is 0.521. In order to future examine every proposed and new dataset can be used.

[4] Harpreeth Kaur J et.al. IDS (Intrusion Detection System) anomaly primarily based totally definitely definitely, study and the conduct of this anomalous via analyzing community webweb page site visitors in awesome standardized datasets. The task that we're confronted in IDS is the huge quantity of facts to method, which results in low detection charge and excessive charge of fake alarm. In this approach, Online Sequential Extreme Learning Machine (OSELM) is proposed for IDS, this method makes use of Symmetric especially relies on feature preference to reduce time complexity

whilst facts beside the point is ignored. Many networks go through threats consisting of spyware, Trojans, viruses, adware, worms, etc. the malwares want to be positioned earlier than any facts lack of tool in enterprise agency or a company. This approach offers IDS that don't overlook approximately severa troubles like feature choice, low accuracy, hugeness of community webweb page site visitors dataset etc. OSELM is a quick and correct Single Hidden Layer and Feed Forward neural community (SHLFH) to method the community.The present tool now not features well and education at the information set is limited.Traits are reused and the trait choice isn't correct withinside the dominant tool. This approach proposed a ultra-contemporary-day method for choice primarily based totally definitely totally on symmetric uncertainty in preference to the CFS approach so we are able to get higher effects. The MLR (multivariate Linear Regression) classifier is used for producing in addition dataset, the accuracy is stepped forward whilst the dataset is trained. The experimental effects provided the classifier detects the assault signatures beautify the accuracy beneath several circumstances. To develop a Network Intrusion Detection System using this way. Machine learning approaches, both supervised and unsupervised, were applied. IDS are trained to detect attack signatures using a variety of datasets, including attack detection. In this concept, multiple ML (Machine Learning) techniques are employed to detect malwares, which modern attackers cannot detect. For malware detection, five machine learning techniques are used: K-Nearest Neighbours, Random Forest, Decision Tree, and Artificial Neural Networks, Logistic Regression. UNSW-NB15 (University of New South Wales) disclosed the dataset used in the system, which contains a considerable quantity of network traffic with many forms of network attacks. The dataset UNSW-NB15 undergoes a number of preparation steps before the class imbalance problem is handled using SMOTE (Synthetic Minority Oversampling Technique) and the accuracy is improvedThe system achieves high accuracy by using SMOET with classifier, which achieved the greatest accuracy of 95.1 percent with a 94.8 percent accuracy rate, 95.7 percent recall, and 95.1 percent F1-score. The precision of the DT classifier was 94.7 percent, which is practically identical to the accuracy of the RF classifier. In terms of accuracy, the RF and DT performance has improved.

3. PROPOSED SYSTEM:

There are several real-lifestyles programs we are the use of nowadays supplied via way of means of system mastering. It seems that control of the system will govern the sector in the coming days. Hence, we got here out right into speculation that the project of figuring out new assaults or zero-day assaults dealing with via way of means of the generation enabled corporations nowadays may be triumph over the use of system mastering methods. Now we evolved a supervised system mastering version that may classifying unseen community visitors primarily based totally on what's learnt from the visible visitors. We use each SVM and

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

ANN algorithm to set of rules to discover the great classifier with better accuracy and fulfilment rate.

To develop Intrusion Detection System will perform the following steps

- Classification
- Building Machine intelligence
- Feature selection

3.1 CLASSIFICATION

Using the training dataset and the features identified in the feature selection section, four models are generated in the Weka software suite. To use supervised machine learning for classification, you must first train the model with a training dataset. As training data, we used 20% of the NSL-KDD dataset, which contains 25,191 labelled data instances. For each sort of feature selection approach, we utilised the SVM and ANN learning algorithms to train the model. As a result, we create four learning models: two SVM models and two ANN models. One of the two models they constructed for each learning method uses 17 features, while the other uses 35 features identified in the feature selection section. Following that, these four trained models were put to the test in 22,542 different ways. The NSL-KDD testing dataset provided 22,542 instances of testing data, which were used to evaluate the models.

3.3.1 Support Vector Machine (SVM)

Support Vector Machine (SVM): In SVM a separating hyper plane defines the classifier depending on the type of problem and available datasets. In case where dataset is one dimensional, the hyper plane is a point, for two-dimensional data it is a separating line as shown in below Figure.

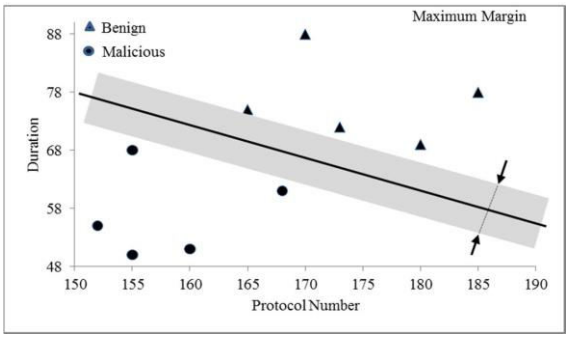

Fig 1: SVM classifier in two dimensional problem spaces

‘

3.1.2 Artificial Neural Network

The Artificial Neural Network (ANN) is a machine learning tool that is almost inspired by the human brain system, and the learning system closely mimics the human brain. A hidden layer exists in both the input and output layers. In ANN, back propagation is the most common method for matching the outcome to the intended result or class.

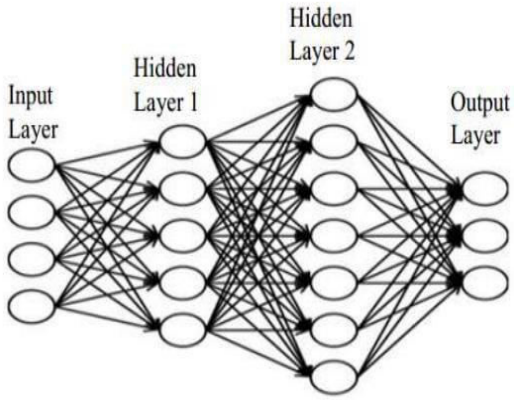

Fig 2: Artificial neural network in the above figure exhibiting input, hidden and output layers

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

3.2 Feature Selection

To reduce data dimensionality, feature selection is a crucial aspect of machine learning, and much research has been done to find a solid feature selection approach. The filter and wrapper methods were used to pick features. Feature scores in various statistical tests that quantify the importance of features by their association with the dependent variable or outcome variable are used to choose features in the filter technique. By evaluating the utility of a subset of features with the dependent variable, the wrapper technique discovers a subset of features. As a result, while filter techniques are independent of any machine learning methodology, the best feature subset selected in the wrapper method is dependent on the machine learning approach used to train the model.

3.3 System Architecture

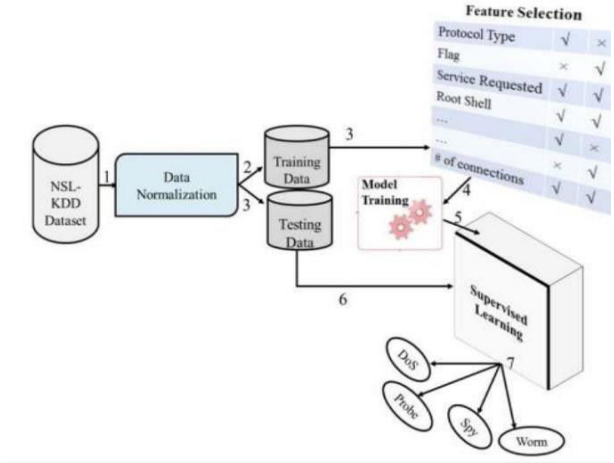

Fig 4: System Architecture

3.4 Dataset Design:

3.4.1 NSL-KDD Data Set

The NSL-KDD dataset is an updated version of the KDD'99 dataset. It has 41 qualities and is classified as either normal or attack, with a single attack type.

Fig 5: NSL-KDD Dataset

3.5 Building Machine Intelligence

Machine learning algorithms are used to develop learning model. Learning models are developed using the best features found in the selection process along with train dataset that is using to train algorithm with the selected features. Each instance has its own training dataset under supervised machine learning.

4. EXPERIMENTAL RESULTS

SVM:

Total Features : 38
Features set reduce after applying features selection concept : 0

Prediction Results

SVM Accuracy, Classification Report & Confusion Matrix

Accuracy : 84.73895582329317

Fig6 : Accuracy of SVM

In above figure we get 84.73 accuracy for SVM.

ANN:

Fig7 : Accuracy of ANN

The obtained accuracy is 96.88% from the above figure.

ACCURACY GRAPH:

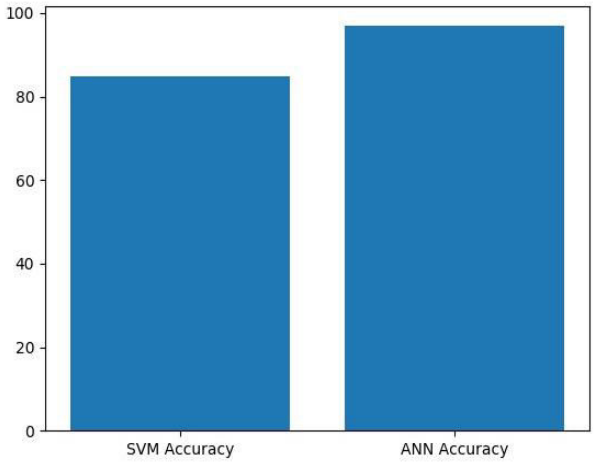

Fig 8: ACCURACY GRAPH

We can determine from the above graph that When comparing ANN to SVM, the x-axis contains the algorithm name and the y-axis reflects the algorithm's accuracy. In the graph above, the x-axis contains the algorithm name and the y-axis shows the algorithm's accuracy.

5. CONCLUSION

In this project, we've got supplied one-of-a-kind device mastering fashions the use of one-of-a-kind device mastering algorithms and one-of-a-kind function choice strategies to discover a excellent version. The evaluation of the end result indicates that the version constructed the use of ANN and wrapper function choice outperformed all different fashions in classifying community site visitors efficaciously with detection fee of 96.88%. We trust that those findings will make contributions to investigate in addition withinside the area of constructing a detection device which could discover recognized assaults in addition to novel assaults. The intrusion detection device exist these days can best discover recognized assaults. Identifying new attacks, often known no attacks, is still a research topic due to the current systems' high fake effective rate.

6. 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 labeled 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, "Anomalybased 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 techniques for intrusion detection," in Communication Networks and Services Research, 2007. CNSR'07. Fifth Annual Conference on, 2007, pp. 350–358.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural networks, vol. 61, pp. 85–117, 2015.

[13] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Military Communications and Information Systems Conference (MilCIS), 2015, 2015, pp. 1–6
.

[14] T. Janarthanan and S. Zargari, "Feature selection in UNSW-NB15 and

KDDCUP'99 datasets," in Industrial Electronics (ISIE), 2017 IEEE 26th International Symposium on, 2017, pp. 1881–1886.

[15] L. Dhanabal and S. P. Shantharajah, "A study on NSL-KDD dataset for intrusion detection system based on classification algorithms," International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 6, pp. 446–452, 2015.