INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 504-508

A NOVEL EFFICIENT INTRUSION DETECTION SYSTEM IN CLOUD USING HYBRID MACHINE LEARNING CLASSIFIER

Arumalla Raja[1], V V Rajesh Babu Anumula[2], Gangolu Rajesh[3], Venigandla Narendra[4], Ashok Kumar Challa[4]

[1,3]Assistant professor, Dept. of ECE, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India.

[2]Assistant professor, Dept. of Library, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India.

[4,5]B. Tech Student, Dept. of ECE, Chalapathi Institute of Technology, Mothadaka, Guntur, Andhra Pradesh, India.

ABSTRACT

Security and Privacy are the biggest issues in widespread cloud systems due to increasing number of Internet-connected devices. A secure cloud system is a major concern for everyone includes government, consumers and business. However attacks on any system are never completely stopped, as a result, real time attacks and threats detection become essential for effective system defense. Intrusion Detection System(IDS) is an enhanced mechanism which is used to control the traffic within the networks and to detect the abnormal activities. Only limited numbers of research works were done on Intrusion Detection Systems (IDS) for Internet of Things (IoT) and cloud. To solve these issues, certain solutions have been designed to improve the security of cloud while monitoring the networks, services and resources and to detect the attacks. On the other hand, Machine Learning techniques are capable for the identification of unknown and known attacks. Over the years, different ML algorithms are used for IDS. However, still there is a lot of scope to achieve better performance for IDS. To fulfill this gap, a novel Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier is presented. The combination of Support Vector Machine (SVM) with Artificial Neural Network (ANN) is presented. The performance of presented IDS is evaluated in terms of Accuracy, F1-score, Recall and Precision.

Keywords: Intrusion Detection, Machine Learning, Support Vector Machine and Artificial Neural Network.

1. INTRODUCTION

One of the world's advance technology is cloud computing. It is an internet-based computer system that provides clients with on-demand access to shared resources such as software, platform, storage, and information. Cloud computing refers to a technology that allows users to access dynamically scaled and virtualized resources via the internet. Cloud computing is one of the latest service innovations in the field of IT [1]. The primary advantage of cloud computing is that it enables access without constraints of location and time. Cloud computing supports mobile and collaborative applications/services, enables the flexibility of controlling storage capacities, and provides lower costs. Moreover, cloud services are multisource, permitting the end-users to use multiple service providers based on their requirements. The use of cloud computing also reduces capital expenditures, power usage, and physical space and maintenance requirements for on-site storage [2].

Customers who use cloud computing don't have to pay for physical infrastructure, which saves money. They rent resources from a third-party supplier, use them as a service, and only pay for the resources they use. Because Small and mid-size enterprises [SMEs] cannot afford the massive capital expenditures required for traditional IT, cloud computing is becoming increasingly linked with them [3].

Cloud computing delivers ubiquitous and pay-per-use services, and as a result of these features, it attracts more consumers to utilize its services. Cloud computing, which is commonly done through the internet, provides services, servers, storage, databases, networking, software, analytics, and more with minimum administration effort. Cloud offers three types of services (IAS, PAS, and SAS) as well as three deployment modes (Private, Public, and Hybrid) to its customers. As a result of these features, the majority of cloud users keep sensitive data on the cloud.

Despite the many benefits of cloud computing, there are a few drawbacks. There is a risk of cloud-based attacks that compromise security features such as confidentiality, availability, and integrity. In order to identify attacks and improve cloud security, security solutions are required for both cloud users and cloud service providers. Cloud computing's primary concern is security [4]. Nowadays, the internet is probably the most available thing around the world. According to internet world stats, there are around 4.57 billion internet users among 7.79 billion people. Online

servers apparently are the store-house of data. Nearly every important information: personal, private, or confidential data are being stored on servers. However, there appears a malicious attack every 39 seconds, which is why about half a billion personal records were stolen by hackers. With the vigorous growth of data available on online servers, the significance of accurate intrusion detection systems is becoming more essential [5].

The increasing prevalence of cyber security attacks has created an imperative for companies to invest in effective tools and techniques for detecting such attacks. Intrusion is a collection of illegal actions that take the system out of safety. The purpose of the intrusion detection is to detect unauthorized use, Abuse and damage to computer systems and networks by both internal users and external attackers. For protecting cloud environments against various threats and cyber attacks, intrusion detection systems (IDS) have become the most extensively used component of computer systems security and security procedures. IDS use a range of reaction mechanisms to discover vulnerabilities, identify illegal activities, and perform prevention measures in order to stay up with the progress of computer-related crimes [6].

Machine learning is one of the world's most

noteworthy techniques to learn from data without explicitly programmed. It's also called the inductive learning method where learner discovers rules by observing examples. Machine learning focuses to provide algorithms that can be trained to perform a task. It involves various methods for analyzing as well assolving classification and regression problems in an effective way. This particular approach is so powerful that it can deal with both labeled (supervised) and unlabeled (unsupervised) data. The reasons to choose machine learning for the detection of intrusion in the first place are pervasive. A machine learning IDS is capable to change its execution strategy as it is acquainted with new information [7].

In this work, a novel Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier is presented. The rest of the work is organized as follows: The section II describes the literature survey. The section III presents Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier. The section IV describes the result analysis. The conclusion is provided in section V.

2. LITERATURE SURVEY

N. Ahmad Hamdi Qaiwmchi, H. Amintoosi and A. Mohajerzadeh et. al., [8] describes Intrusion Detection System Based on Gradient Corrected Online Sequential Extreme Learning Machine. This article proposes a novel approach for finding the optimal weights of the input-hidden layer. This article presents an approach for an integration between OSELM and back-propagation designated as (OSELM-BP). After integration, BP changes the random weights iteratively and uses an iterated evaluation of the generated error for feedback correction of the weights. The approach is evaluated based on various scenarios of activation functions for OSELM on the one hand and the number of iterations for BP on the other. An extensive evaluation of the approach and comparison with the original OSELM reveal a superiority of OSELM-BP in reaching optimal accuracy with a small number of iterations. However, a degradation in the performance happens when the number of neurons is higher due to over-fitting.

B. A. Tama, M. Comuzzi and K. -H. Rhee et. al., [9] describes TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System. an improved IDS based on hybrid feature selection and two-level classifier ensembles are proposed. A hybrid feature selection technique comprising three methods, i.e., particle swarm optimization, ant colony algorithm, and genetic algorithm, is utilized to reduce the feature size of the training datasets (NSL-KDD and UNSW-NB15 are considered in this paper). Features are selected based on the classification performance of a reduced error pruning tree (REPT) classifier. Then, a two-level classifier ensemble based on two meta learners, i.e., rotation forest and bagging, is proposed. On the NSL-KDD dataset, the proposed classifier shows 85.8% accuracy, 86.8% sensitivity, and 88.0% detection rate, which remarkably outperform other classification techniques recently proposed in the literature. It is required to validate the proposed approach in solving a multi-class classification problem, which represents incoming network traffic as normal or some attack groups.

A. Javadpour, S. Kazemi Abharian and G. Wang et. al., [10] describes Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms. The purpose of this study is to provide a new method based on intrusion detection systems and its various architectures aimed at increasing the accuracy of intrusion detection in cloud computing. Different classification algorithms including decision tree, random forest, CART algorithm and neural network were applied to the data. Combination information methods such as majority voting can also be used to improve the results achieved in classifications. S. Lv, J. Wang, Y. Yang and J. Liu et. al., [11] presents Intrusion Prediction With System-Call Sequence-to-Sequence Model. This prediction model predicts a sequence of

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 504-508

e-ISSN : 2583-1062

Impact Factor : 5.725

system-calls that will be executed in the future, which will enable the monitoring of system state and the prediction of attack behavior. The experiments show that the prediction method proposed in this paper achieved well prediction performance on ADFA-LD intrusion detection test data set. Moreover, the predicted sequence, combined with the known invoked traces of system call, significantly improves the performance of intrusion detection verified on various classifiers. Further work is needed to be done to enhance the model robustness against adversarial samples.

M. E. KarsligE, A. G. Yavuz, M. A. Güvensan, K. Hanifi and H. Bank et. al., [12] presents Network intrusion detection using machine learning anomaly detection algorithms. A NSL-KDD labelled dataset of network connection traces is used -for testing our method's effectiveness. The experiments result on the NSL-KDD data set, shows that we achieved an accuracy of 80.119%.

Samir Ifzarne, Hiba Tabbaa, Imad Hafidi, Nidal Lamghari et. al., [2020] describes Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks. This model is built based on information gain ratio and the online Passive aggressive classifier. Firstly, the information gain ratio is used to select the relevant features of the sensor data. Secondly, the online Passive aggressive algorithm is trained to detect and classify different type of Deny of Service attacks. The experiment was conducted on a wireless sensor network-detection system (WSN-DS) dataset. The proposed model ID-GOPA results detection rate of 96% determining whether the network is in its normal mode or exposed to any type of attack. The detection accuracy is 86%, 68%, 63%, and 46% for scheduling, grayhole, flooding and blackhole attacks, respectively.

A NOVEL EFFICIENT INTRUSION DETECTION SYSTEM IN CLOUD

In this section, a novel Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier is presented. The block diagram of presented IDS is shown in Figure 1.

In order to effectively train machine learning algorithms, a significant quantity of data is necessary. The quality and quantity of data are crucial factors in every machine learning evaluation. Undoubtedly, these concerns are significantly reliant on data: higher quality data often exceeds more efficient algorithms, which are of crucial significance.

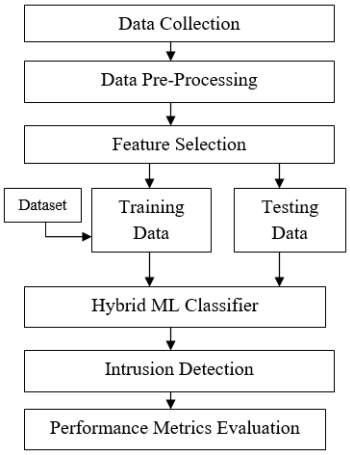

Figure 1: Block Diagram of novel Efficient Intrusion detection system in cloud

The dataset CSE-CIC-IDS2018 (Communications Security Establishment- Canadian Institute for Cyber security-Intrusion Detection System) is used in this study which is introduced by the "Canadian Institute for Cyber security (CIC)" which is recent and is the most realistic cyber dataset in 2018. The CIC-IDS2018 dataset is well-suited for classification and detecting time series anomalies in cloud computing environments using machine learning techniques. This is due to its realistic and comprehensive data, established benchmark status, availability of labeled data, and dynamic and diverse characteristics. The dataset contains diverse network traffic data, including various types of attacks commonly found in real-world cloud computing environments, making it a reliable choice for evaluating intrusion detection methods.

The CICIDS2018 dataset includes captures and arrangements of machine traffic and framework logs, as well as 80 features selected from the traffic captured using the CICFlowMeter-V3 traffic flow generator. The dataset is in CSV format, with six main features labeled as SourceIP, SourcePort, DestinationIP, DestinationPort, FlowID, and 80 attributes labeled as Protocol.

Prior to dataset processing, a pre-processing step is necessary to adequately handle the presence of both numerical and non-numerical instances. Data cannot be fed to most machine learning classifiers if any of the features contain non-numeric values. The preprocessing module focused on data normalization. Therefore, the categorical features are

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 504-508

transformed into numeric values with the dummies function that allows symbolic features to be mapped as numeric values. Then, the detected inconsistencies are deleted.

A feature selection technique can be seen as a procedure for selecting a precise, compact and accurate subset of features from a given feature set. Features selection (FS) and dimension reduction methods are important in machine learning for optimizing prediction inputs and improving prediction accuracy. FS involves selecting relevant features based on selection and stopping criteria, and results are evaluated using evaluation criteria. The objective is to select a limited number of features to categorize network data, which can not only reduce training time but also improve accuracy by removing irrelevant information.

The process of partitioning the complete dataset into two distinct subsets, namely test and train data where 80% of data is used to train the Hybrid ML model and a 20% data is used to test the data. Testing and training data is applied to hybrid ML classifier. The support vector machine is a machine learning technique known as the best learning algorithm for classification. It is one of the most popular supervised ML algorithms. The SVM is a type of pattern classification and regression with a variety of kernel functions. It has been applied to several pattern recognition applications. SVMs have been used mainly for binary classification. The idea is to find a line or hyperplane that separates two classes such that it is as far as possible from the closest instances of each class. Instances of different groups are separated by an area called the margin. The closest instances to the hyperplane are called support vectors. It is required to have the margin as big as possible; this is important to enhance the efficiency of the classification of newly added instances. The support vector machine can determine the appropriate setting parameters because it does not depend on traditional experimental risks and can learn a wide range of patterns that can expand. Support vector machines can also dynamically update training patterns whenever there is a new pattern during classification.

Artificial neural networks (ANNs), also shortened to neural networks (NNs) or *neural nets*) are a branch of machine learning models that are built using principles of neuronal organization discovered by connectionism in the biological neural networks constituting animal brains. ANNs make use of a “learning rule” (often gradient descent based back-propagation of errors) that allows the set of weights and biases for the hidden layer and output layer neurons to be adaptively tuned. This self-adaptive nature means that ANNs are capable of capturing highly complex and non-linear relationships between both dependent and independent variables without prior knowledge.

The Hybrid ML classifier is used to detect the anomalies, attacks and intruders in cloud.s The primary aim of an Intrusion Detection System (IDS) is to identify when a malefactor is attempting to compromise the

operation of a system. This research offers a hybrid model system that uses two machine learning techniques. In order to provide the highest level of intrusion detection in the cloud, a hybrid method is used. The proposed hybrid model employs the supervised learning algorithm ANN and SVM. The process of evaluation involves providing the test data into the model and comparing the expected attack categories with the accurate labels.

3. RESULT ANALYSIS

In this section, a novel Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier is presented. The result analysis of presented system is demonstrated here. Various metrics are employed to characterize the effectiveness of presented approach and are defined as follows:

Recall: It refers to the ratio of accurately identified attacks to the total number of attacks.

Precision refers to the ratio of correctly identified attacks to the total number of identified attacks.

Accuracy: Accuracy is a metric used to measure the proportion of correctly identified outcomes in both normal and attack traffic.

F1-score: The F1-score is a statistical measure that represents the harmonic mean of precision and recall.

table 1 validates the performance analysis.

Table 1: Performance Validation

Metrics/Classifier	Naïve Bayes (NB)	Hybrid ML(HML) classifier
Recall (%)	86	96.25
Precision (%)	88	97.6
Accuracy (%)	85.3	96.54
F1-score	82.3	96.7

Compared to NB classifier, presented HML classifier has better performance. The Figure 2 shows recall and Precision performance comparison.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

e-ISSN : 2583-1062

Impact Factor : 5.725

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 01, January 2024, pp : 504-508

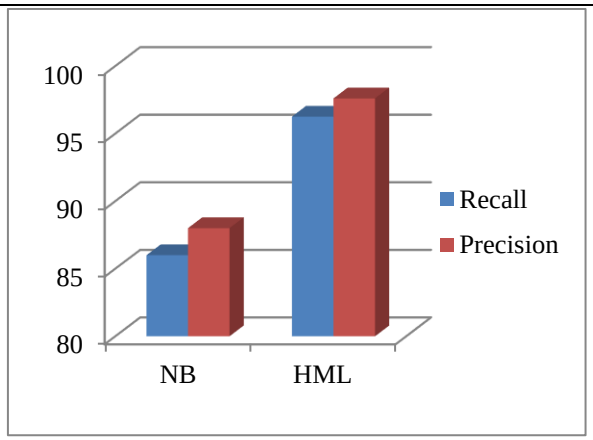

Figure 2: Performance Comparison

In figure 2, the x-axis indicates classifiers and y-axis indicates performance comparison. The Hybrid ML classifier has high recall and precision than NB classifier. The Figure 3 shows Accuracy and F1-score performance comparison.

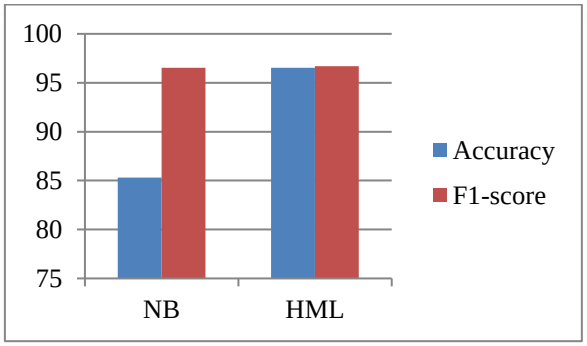

Figure 3: Accuracy and F1-score Comparison

Compared to NB classifier, the Hybrid ML (HML) classifier has high accuracy and F1-score. This system has detected the attacks and intruders very effectively compared to previous systems.

4. CONCLUSION

In this work, a novel Efficient Intrusion detection system in cloud using Hybrid Machine Learning classifier is presented. CIC-IDS2018 dataset is used which includes captures and arrangements of machine traffic and framework logs, as well as 80 features selected from the traffic captured using the CICFlowMeter-V3 traffic flow generator. The dataset is divided into 80% training and 20% is used to test the data. The ANN and SVM classifiers are combined as a Hybrid ML classifier to detect the attacks and intruders. The performance of presented IDS is validated in terms of Accuracy, Precision, Recall and F1-score. Compared to previous classifiers, this Hybrid Machine Learning classifier has better performance in terms of Accuracy, Precision, Recall and F1-score. This system has detected various attacks and intruders very accurately and efficiently.

5. REFERENCES

[1] H. Attou, A. Guezzaz, S. Benkirane, M. Azrour and Y. Farhaoui, "Cloud-Based Intrusion Detection Approach Using Machine Learning Techniques," in Big Data Mining and Analytics, vol. 6, no. 3, pp. 311-320, September 2023, doi: 10.26599/BDMA.2022.9020038.

[2] Abdel-Rahman Al-Ghuwairi, Yousef Sharrab, Dimah Al-Fraihat, Majed AlElaimat, Ayoub Alsarhan and Abdulmohsen Algarni, “Intrusion detection in cloud computing based on time series anomalies utilizing machine learning,” Journal of Cloud Computing, vol. 12, no. 127,doi:10.1186/s13677-023-00491-x

[3] Ammar Aldallal and Faisal Alisa, “Effective Intrusion Detection System to Secure Data in Cloud Using Machine Learning,” Symmetry, vol. 13, no. 12, pp. 1-26, 2021, doi:10.3390/sym13122306

[4] Alshahrani, Hani, Attiya Khan, Muhammad Rizwan, Mana Saleh Al Reshan, Adel Sulaiman, and Asadullah Shaikh, "Intrusion Detection Framework for Industrial Internet of Things Using Software Defined Network," Sustainability, vol. 15, no. 11, 2023, doi:10.3390/su15119001

[5] G. Abdelmoumin, D. B. Rawat and A. Rahman, "On the Performance of Machine Learning Models for Anomaly-Based Intelligent Intrusion Detection Systems for the Internet of Things," in IEEE Internet of Things Journal, vol. 9, no. 6, pp. 4280-4290, 15 March15, 2022, doi: 10.1109/JIOT.2021.3103829.

[6] Khalid Al Makdi, Frederick T Sheldon and Soule Terence, "Adaptive Trust-based Security Model for Intrusion Detection Using Deep Learning Technique in the Cloud," Acta Scientific COMPUTER SCIENCES, Volume 4 Issue 12 December 2022,

[7] Md. Badiuzzaman Pranto, Md. Hasibul Alam Ratul, Md. Mahidur Rahman, Ishrat Jahan Diya, and Zunayeed-Bin Zahir, "Performance of Machine Learning Techniques in Anomaly Detection with Basic Feature Selection Strategy - A Network Intrusion Detection System," Journal of Advances in Information Technology Vol. 13, No. 1, February 2022, doi: 10.12720/jait.13.1.36-44

[8] N. Ahmad Hamdi Qaiwmchi, H. Amintoosi and A. Mohajerzadeh, "Intrusion Detection System Based on Gradient Corrected Online Sequential Extreme Learning Machine," in IEEE Access, vol. 9, pp. 4983-4999, 2021, doi: 10.1109/ACCESS.2020.3047933.

[9] B. A. Tama, M. Comuzzi and K. -H. Rhee, "TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System," in IEEE Access, vol. 7, pp. 94497-94507, 2019, doi: 10.1109/ACCESS.2019.2928048.

[10] A. Javadpour, S. Kazemi Abharian and G. Wang, "Feature Selection and Intrusion Detection in Cloud Environment Based on Machine Learning Algorithms," 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), Guangzhou, China, 2017, pp. 1417-1421, doi: 10.1109/ISPA/IUCC.2017.00215.

[11] S. Lv, J. Wang, Y. Yang and J. Liu, "Intrusion Prediction With System-Call Sequence-to-Sequence Model," in IEEE Access, vol. 6, pp. 71413-71421, 2018, doi: 10.1109/ACCESS.2018.2881561.

[12] M. E. KarsligE, A. G. Yavuz, M. A. Güvensan, K. Hanifi and H. Bank, "Network intrusion detection using machine learning anomaly detection algorithms," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, 2017, pp. 1-4, doi: 10.1109/SIU.2017.7960616.

[13] Samir Ifzarne, Hiba Tabbaa, Imad Hafidi, Nidal Lamghari, "Anomaly Detection using Machine Learning Techniques in Wireless Sensor Networks," Journal of Physics: Conference Series, vol. 1743, pp. 1-23, 2021, doi:10.1088/1742-6596/1743/1/012021

[14] I. Aljamal, A. Tekeoğlu, K. Bekiroglu and S. Sengupta, "Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environments," 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), Honolulu, HI, USA, 2019, pp. 84-89, doi: 10.1109/SERA.2019.8886794.

[15] Dhivya R, Dharshana R, Divya V, "Security Attacks Detection in Cloud using Machine Learning Algorithms," International Research Journal of Engineering and Technology (IRJET), vol. 06, no. 02, Feb 2019, e-ISSN: 2395-0056