IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

ANOMALIES INTRUSION DETECTION IN CLOUD COMPUTING USING MACHINE LEARNING

[1]Dr. R. Rambabu, [2]Mr. Ch. Suryababu, [3]Dr. M. Sridhar

[1]Professor & HOD, Department of Computer Science and Engineering, Rajamahendri Institute of Engineering and Technology, Bhoopalapatnam, Rajamahendravaram, Andhra Pradesh, India

[2]Associate Professor, Department of Electronics and Communication Engineering, Adarsh College of Engineering, Andhra Pradesh, India

[3]Principal, Bharath Institute of Engineering and Technology, Ibrahimpatnam (M), Andhra Pradesh, India

Abstract: With the rapid adoption of cloud computing, ensuring the security of sensitive data and resources has become a major concern. Traditional security methods often fall short in the dynamic and distributed nature of cloud environments. This paper presents an approach for detecting anomalies and intrusions in cloud computing using machine learning techniques. By using CSE-CIC IDS 2018 Dataset the system is able to identify abnormal behaviors and malicious activities that may indicate security threats. The proposed method analyzes cloud traffic, system logs, and user behavior to uncover potential intrusions, with the goal of enhancing the overall security posture of cloud systems. Experimental results show that machine learning-based anomaly detection offers high detection accuracy and low false positive rates, making it an effective tool for proactive security in cloud environments.

Keywords: Cloud Computing, Anomalies Detection, Machine Learning Techniques, CSE-CIC IDS 2018 Dataset, Accuracy, Low False Positive Rates.

I. INTRODUCTION

Cloud computing has revolutionized the way businesses and individuals manage their data, applications and services. By offering scalable, on-demand resources, cloud platforms have significantly reduced the cost of infrastructure while providing flexibility and high availability. However, the rapid adoption of cloud services has introduced new security challenges. Cloud environments are basically danger due to their multi-tenant nature, dynamic workloads and vast attack surface. Intrusions and anomalies in cloud systems can lead to data breaches, service disruptions, and financial losses, making intrusion detection a critical component of cloud security. Traditional security mechanisms, including firewalls and signature-based intrusion detection systems (IDS), often fall short in detecting novel or sophisticated attacks due to their reliance on predefined patterns.

Machine learning (ML) techniques offer a promising solution for addressing the limitations of conventional security tools. ML algorithms can automatically learn and adapt to new patterns in data, making them well-suited for detecting complex and previously unseen threats in cloud environments. By training models on large datasets, these systems can identify anomalous behavior, which might indicate an intrusion, without requiring explicit programming for every possible attack. The

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

ability of machine learning to generalize from data and its capacity for real-time detection make it an ideal candidate for intrusion detection in cloud computing environments.

The CSE-CIC IDS 2018 dataset provides a valuable resource for evaluating machine learning-based intrusion detection models. This dataset, developed by the Canadian Institute for Cyber security, contains a comprehensive collection of network traffic data, including normal traffic as well as various types of attacks, such as Denial of Service (DoS), Brute Force and Botnet attacks. It consists of both labeled and unlabeled data, offering an opportunity to train and test supervised and unsupervised machine learning models. The diversity of attack types and the realistic traffic patterns make the CSE-CIC IDS 2018 dataset ideal for developing and testing anomaly detection systems that can effectively operate in cloud environments.

Anomaly detection is particularly useful in cloud security because it focuses on identifying deviations from normal patterns, which can be indicative of an intruder or malicious activity. Machine learning models can be trained on normal system behavior, and once deployed, they continuously monitor real-time cloud traffic to flag unusual patterns. This ability to detect unknown or novel attacks without relying solely on previously identified signatures provides a significant advantage over traditional methods. However, the challenge lies in efficiently training models to distinguish between benign anomalies, such as system misconfigurations, and harmful intrusions, which requires large and varied datasets for effective learning.

This paper proposes an anomaly-based intrusion detection system (IDS) for cloud computing environments, using machine learning techniques with the CSE-CIC IDS 2018 dataset. Our objective is to develop a robust and scalable solution for detecting intrusions in the cloud by leveraging the power of machine learning algorithms. We aim to demonstrate the feasibility and effectiveness of applying these models to real-world cloud traffic data, highlighting their potential to enhance the security and toughness of cloud systems. By evaluating multiple machine learning techniques, we seek to identify the most suitable methods for anomaly detection in this evolving and high-risk domain.

II. LITERATURE SURVEY

Z. Chkirbene, R. Hamila, A. Erbad, S. Kiranyaz, N. Al-Emadi and M. Hamdi et al. Cloud computing is attracting a lot of attention in the past few years. Although, even with its wide acceptance, cloud security is still one of the most essential concerns of cloud computing. Many systems have been proposed to protect the cloud from attacks using attack signatures. Most of them may seem effective and efficient; however, there are many drawbacks such as the attack detection performance and the system maintenance. Recently, learning-based methods for security applications have been proposed for cloud anomaly detection especially with the advents of machine learning techniques. However, most researchers do not consider the attack classification which is an important

parameter for proposing an appropriate countermeasure for each attack type. In this paper, we propose a new firewall model called Secure Packet Classifier (SPC) for cloud anomalies detection and classification. The proposed model is constructed based on collaborative filtering using two machine learning algorithms to gain the advantages of both learning schemes. This strategy increases the learning performance and the system's accuracy. To generate our results, a publicly available dataset is used for training and testing the performance of the proposed SPC. Our results show that the accuracy of the SPC model increases the detection accuracy by 20% compared to the existing machine learning algorithms while keeping a high attack detection rate [1].

I. Aljamal, A. Tekeoğlu, K. Bekiroglu and S. Sengupta et al. Intrusion detection is one essential tool towards building secure and trustworthy Cloud computing environment, given the ubiquitous presence of cyber attacks that proliferate rapidly and morph dynamically. In our current working paradigm of resource, platform and service consolidations, Cloud Computing provides a significant improvement in the cost metrics via dynamic provisioning of IT services. Since almost all cloud computing networks lean on providing their services through Internet, they are prone to experience variety of security issues. Therefore, in cloud environments, it is necessary to deploy an Intrusion Detection System (IDS) to detect new and unknown attacks in addition to signature based known attacks, with high accuracy. In our deliberation we assume that a system or a network “anomalous” event is synonymous to an “intrusion” event when there is a significant departure in one or more underlying system or network activities. There are couple of recently proposed ideas that aim to develop a hybrid detection mechanism, combining advantages of signature-based detection schemes with the ability to detect unknown attacks based on anomalies. In this work, we propose a network based anomaly detection system at the Cloud Hypervisor level that utilizes a hybrid algorithm: a combination of K-means clustering algorithm and SVM classification algorithm, to improve the accuracy of the anomaly detection system. Dataset from UNSW-NB15 study is used to evaluate the proposed approach and results are compared with previous studies. The accuracy for our proposed K-means clustering model is slightly higher than others. However, the accuracy we obtained from the SVM model is still low for supervised techniques [2].

A. R. Wani, Q. P. Rana, U. Saxena and N. Pandey et al. The primary benefit of the cloud is that it elastically scales to meet variable demand and it provides the environment which scales up and scales down instantly according to the demand, so it needs great protection from DDoS attacks to tackle downtime effects of DDoS Attacks. Distribute Denial of Service attacks fall on the category of critical attacks that compromise the availability of the network. These attacks have become sophisticated and continue to grow at a rapid pace so to detect and counter these attacks have become a challenging task. This work was carried out on the owncloud environment using Tor Hammer as an attacking tool and a new dataset was generated with Intrusion Detection System. This work incorporates

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

various machine learning algorithms: Support Vector Machine, Naive Bayes, and Random Forest for classification and overall accuracy was 99.7%, 97.6% and 98.0% of Support Vector Machine, Random Forest and Naive Bayes respectively [3].

H. Yang et al. With the continuous expansion of cloud computing, the intrusion of civil aviation internal network and the abnormal behavior of internal users have become a new problem of cloud computing security. Aiming at the problem of intrusion detection and abnormal behavior analysis of internal network, this paper studies the intrusion detection data set by using the machine learning typical classified algorithm of Weka software, and realizes the malicious behavior of civil aviation internal network users through naive Bayesian algorithm, the behavior and normal behavior are classified and analyzed, and the naive Bayesian algorithm has high classified accuracy for abnormal behavior. It can effectively classify and mine the internal network user behavior of cloud computing intrusion detection data. Through experiments, the effectiveness of the proposed scheme and algorithm is verified [4].

Q. Schueller, K. Basu, M. Younas, M. Patel and F. Ball et al. Software-Defined Networks (SDN) has emerged as a dominant programmable network architecture for cloud based data centers. Its centralised programmable control plane decoupled from the data plane with a global view of the network state provides new opportunities to implement innovate security mechanisms. This research Ieverages this features of SDN and presents the architecture of a hierarchical and light weight Intrusion Detection System (IDS) for software enabled networks by exploiting the concept of SDN flows. It combines advantages of a flow-based IDS and a packet-based IDS in order to provide a high detection rate without degrading network performances. The flow-based IDS uses an anomaly detection algorithm based on Support Vector Machines (SVM) trained with DARPA Intrusion Detection Dataset. This first line of defence detects any intrusions on the network. When an attack is detected, the malicious flow is mirrored to a packet-based IDS, for further examination and actions. The results show that this scheme provides good detection rates and performances with minimal extra overhead [5].

III. METHODOLOGY

The anomaly-based intrusion detection in cloud computing using machine learning involves several key stages: data preprocessing, feature extraction, feature selection, training and evaluation, and system deployment. Each stage plays a critical role in ensuring the effectiveness and accuracy of the proposed anomaly detection system. In this section, we outline the steps taken to implement a machine learning-driven intrusion detection system using the CSE-CIC IDS 2018 dataset.

The CSE-CIC IDS 2018 dataset consists of network traffic data, including both normal and attack scenarios. The dataset includes a variety of attack types, such as Denial of Service (DoS), Distributed Denial of Service (DDoS), Brute Force, and Botnet attacks, among others. Since the data is highly complex and may contain redundant or

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

irrelevant information, the first step in the methodology is to preprocess the data. This involves data cleaning, handling missing values, and removing duplicates. Normalization is applied to scale features to a standard range, ensuring that no feature dominates the learning process. Additionally, categorical variables are encoded into numerical representations, and time stamp-based features are extracted for temporal analysis. This step ensures that the dataset is in a format suitable for training machine learning models.

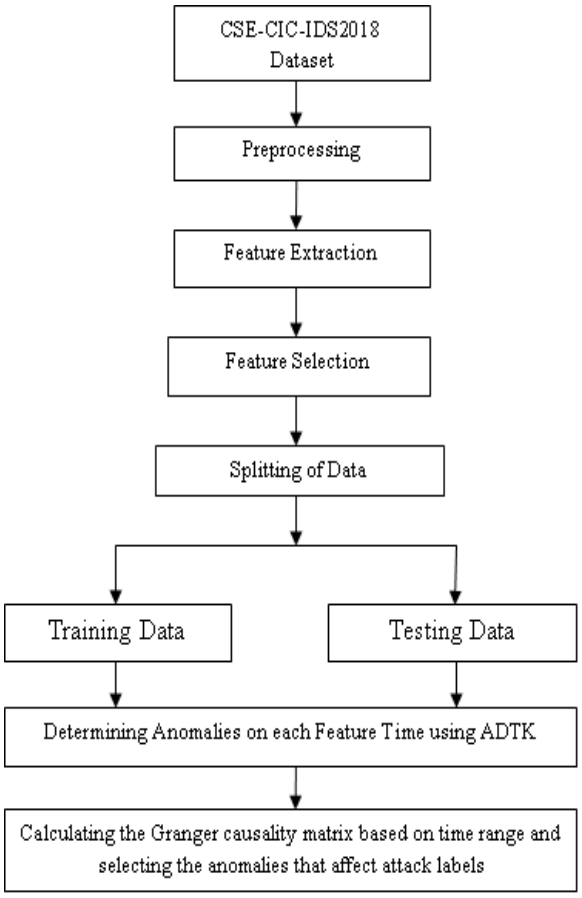

Fig. 1: Block Diagram

Feature extraction plays a crucial role in identifying patterns in network traffic that can indicate anomalous behavior. Raw traffic data, such as packet sizes, time stamps and protocol types, are insufficient by themselves for accurate intrusion detection. Therefore, we derive additional features that reflect higher-level network behavior, such as the number of connections, session durations and request/response ratios. Aggregating traffic statistics over different time windows is essential for capturing both short-term and long-term trends in the network. Furthermore, statistical features like mean, variance and standard deviation are computed for various traffic metrics. These engineered features improve the model's ability to differentiate between normal operations and attack patterns.

Once the data is preprocessed and relevant features are extracted, the next step is to choose appropriate machine learning models for anomaly detection. The next phase involves training the selected machine learning models using a portion of the CSE-CIC IDS 2018 dataset. A training and testing split is applied to avoid over fitting and ensure that the model generalizes well to unseen data. Cross-validation is used to fine-tune hyper parameters and select the optimal model configuration. To evaluate the performance of the models, several metrics are employed, including accuracy, precision, recall, F1-score,and the area under the Receiver Operating Characteristic (ROC) curve. Precision and recall are particularly important in the context of intrusion detection, as we aim to minimize both false positives (benign activities misclassified as attacks) and false negatives (missed attacks). The evaluation also involves analyzing the detection rates of different attack types and

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

assessing how well the model can identify novel or previously unseen threats.

After model evaluation, the best-performing machine learning model is selected for real-time deployment in a cloud environment. The model is integrated into the cloud system's security framework, where it continuously monitors incoming traffic and analyzes behavior in real-time. In practice, the system will flag any traffic that deviates significantly from learned normal behavior as anomalous. Once an anomaly is detected, the system triggers an alert to the administrators for further investigation, allowing for swift mitigation of potential security breaches.

To ensure continuous improvement, the system is designed to adapt over time by retraining the model with new data, capturing evolving patterns of network behavior and emerging attack vectors. Additionally, the system is equipped with feedback mechanisms that allow it to refine its detection capabilities based on user input and newly discovered attack techniques.

By following this methodology, the proposed anomaly-based intrusion detection system not only leverages machine learning techniques to enhance detection accuracy but also adapts to the dynamic and evolving nature of cloud computing environments, providing a robust defense against a wide range of cyber threats.

IV. RESULTS

In this section performance analysis of anomalies intrusion detection in cloud computing using machine learning is observed.

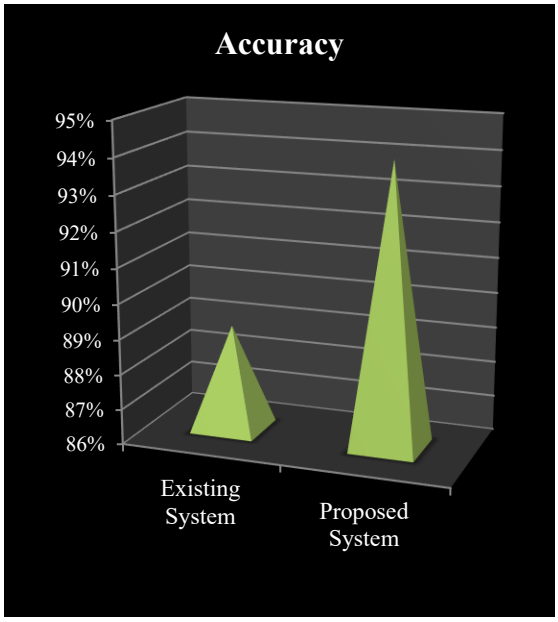

Fig. 2: Accuracy Comparison Graph

In figure 2 shows the accuracy comparison graph is observed between existing system and proposed system. The proposed system shows the high accuracy.

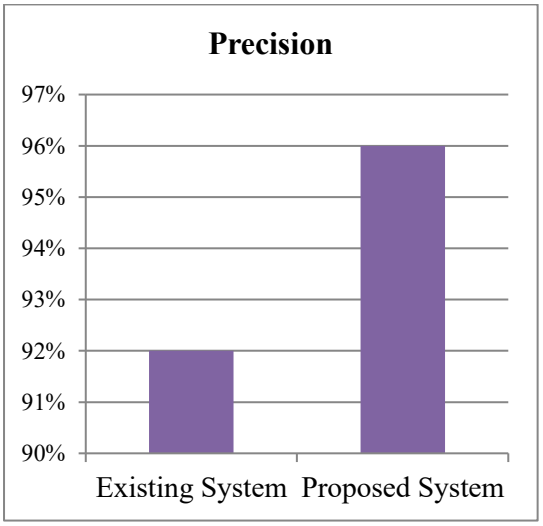

Fig. 3: Precision Comparison Graph

In figure 3 shows the precision comparison graph is observed between existing system and proposed system. The proposed system shows the better precision.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11, Iss 13, 2022

V. CONCLUSION

This paper demonstrates the effectiveness of machine learning-based anomaly detection for securing cloud environments. By leveraging the CSE-CIC IDS 2018 dataset, the proposed system successfully identifies abnormal behaviors and malicious activities, offering high detection accuracy and low false positive rates. This approach enhances the security posture of cloud systems, providing a proactive and adaptive defense mechanism against a wide range of cyber threats. Hence this paper achieves better results in terms of accuracy and precision.

VI. REFERENCES

[1] Z. Chkirbene, R. Hamila, A. Erbad, S. Kiranyaz, N. Al-Emadi and M. Hamdi, "Cooperative Machine Learning Techniques for Cloud Intrusion Detection," 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin City, China, 2021, pp. 837-842, doi: 10.1109/IWCMC51323.2021.9498809.

[2] 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.

[3] A. R. Wani, Q. P. Rana, U. Saxena and N. Pandey, "Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 870-875, doi: 10.1109/AICAI. 2019.8701238.

[4] H. Yang, "Research on Classification Algorithm for Civil Aviation Internal Network Intrusion Detection Based on Machine Learning," 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, Weihai, China, 2020, pp. 1-4, doi: 10.1109/ICCASIT50869.2020.9368594.

[5] Q. Schueller, K. Basu, M. Younas, M. Patel and F. Ball, "A Hierarchical Intrusion Detection System using Support Vector Machine for SDN Network in Cloud Data Center," 2018 28th International Telecommunication Networks and Applications Conference (ITNAC), Sydney, NSW, Australia, 2018, pp. 1-6, doi: 10.1109/ATNAC.2018.8615255.

[6] Zina Chkirbene, Sebti Foufou, Ridha Hamila, Zahir Tari and Albert Y. Zomaya, "Lacoda: Layered connected topology for massive data centers", Journal of Network and Computer Applications, vol. 83, pp. 169-180, 2017.

[7] Lav Gupta, Raj Jain, Mohammed Samaka, Aiman Erbad and Deval Bhamare, "Performance evaluation of multi-cloud management and control systems", Recent Advances in Communications and Networking Technology (Formerly Recent Patents on Telecommunication), vol. 5, no. 1, pp. 9-18, 2016.

[8] Muhammad Abedin, Syeda Nessa, Latifur Khan and Bhavani Thuraisingham, "Detection and resolution of anomalies in

firewall policy rules" in Data and Applications Security XX, Berlin, Heidelberg:Springer Berlin Heidelberg, pp. 15-29, 2006.

[9] Deval Bhamare, Maede Zolanvari, Aiman Erbad, Raj Jain, Khaled Khan and Nader Meskin, "Cybersecurity for industrial control systems: A survey", computers & security, vol. 89, pp. 101677, 2020.

[10] Zina Chkirbene, Aiman Erbad, Ridha Hamila, Ala Gouissem, Amr Mohamed, Mohsen Guizani, et al., "Weighted trustworthiness for ml based attacks classification", 2020 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-7, 2020.

[11] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems (UNSW-NB15 Network Data Set)", 2015 Military Communications and Information Systems Conference (MilCIS), pp. 1-6, Nov 2015.

[12] S. Aljawarneh, M. Aldwairi and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model", Journal of Computational Science, vol. 25, pp. 152-160, 2018.

[13] S. Baek, D. Kwon, J. Kim, S. C. Suh, H. Kim and I. Kim, "Unsupervised Labeling for Supervised Anomaly Detection in Enterprise and Cloud Networks", 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CSCloud), pp. 205-210, June 2017.

[14] A. Aldribi, I. Traore and B. Moa, "Hypervisor-Based Cloud Intrusion Detection through Online Multivariate Statistical Change Tracking", IEEE Transactions on Information Forensics and Security, 2018.

[15] N. Krishnan and A. Salim, "Machine Learning Based Intrusion Detection for Virtualized Infrastructures," 2018 International CET Conference on Control, Communication, and Computing (IC4), Thiruvananthapuram, India, 2018, pp. 366-371, doi: 10.1109/CETIC4.2018.8530912.

[16] W. Haider, J. Hu, Y. Xie, X. Yu and Q. Wu, "Detecting anomalous behavior in cloud servers by nested-arc hidden semi-markov model with state summarization", IEEE Transactions on Big Data, vol. 14, no. 8, Aug. 2015.

[17] D. Sun, M. Fu, L. Zhu, G. Li and Q. Lu, "Non-intrusive anomaly detection with streaming performance metries and logs for DevOps in public clouds: A case study in AWS", IEEE Transactions on Emerging Topics in Computing, vol. 4, no. 2, pp. 278-289, 2016.

[18] J. Zhang, Y. Xiang, Y. Wang, W. Zhou, Yo. Xiang and Y. Guan, "Network traffic classification using correlation information", IEEE Transactions on Parallel and Distributed Systems, vol. 24, no. 1, Jan. 2013.