ADAPTIVE INTRUSION DETECTION SYSTEMS USING MACHINE LEARNING IN CLOUD ENVIRONMENTS

Ravindrakumar [1]

[1] Assistant Professor, Department of Computer Science, Government First Grade College Chitaguppa

DOI
10.29121/shodhkosh.v4.i1.2023.3358

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

There are considerable hurdles that must be overcome in order to maintain adequate security against cyber-attacks in cloud settings because of their dynamic and dispersed nature. A proactive method to identifying and mitigating new threats in real time is provided by adaptive intrusion detection systems (IDS) that make use of machine learning (ML). The design and operation of adaptive intrusion detection systems (IDS) that are adapted for cloud platforms are investigated in this study. Particular attention is paid to the utilisation of supervised, unsupervised, and reinforcement learning methods. It investigates the benefits of adaptive intrusion detection systems (IDS) in terms of identifying abnormal behaviours, minimising the number of false positives, and adapting to shifting threat environments. Several important factors, including the selection of features, the quality of the dataset, and the optimisation of the algorithm, are covered. In addition to this, we investigate the possibility of integrating adaptive intrusion detection systems with cloud-native technologies such as serverless computing and containers. Performance measurements and comparative evaluations demonstrate that machine learning-based intrusion detection systems are more effective than older techniques. Additionally, the study addresses issues such as scalability, data privacy, and adversarial assaults, and it proposes viable methods to improve dependability. The adaptive intrusion detection system (IDS) is an essential component of cloud security methods since it enables continuous monitoring and reaction to complex kinds of attacks.

Keywords: Intrusion Detection Systems, Machine Learning, Cloud Security, Adaptive Systems, Anomaly Detection, Real-Time Threat Mitigation, Adversarial Attacks

1. INTRODUCTION

Cloud computing has been rapidly adopted, which has resulted in a transformation of the digital environment. Cloud computing offers unparalleled flexibility, scalability, and cost effectiveness. On the other hand, these benefits are accompanied by substantial issues about security. Due to the fact that cloud environments are dynamic and dispersed, they are vulnerable to a wide variety of sophisticated cyber attacks, which compels them to implement stringent security measures. Among these precautions, Intrusion Detection Systems (IDS) have emerged as indispensable instruments for the protection of cloud infrastructures. As a result of the incorporation of machine learning (ML) techniques, intrusion detection systems (IDS) have developed into more adaptable systems that are able to recognise intricate attack patterns, making them a vital component of contemporary cloud security.

How to cite this article (APA): Ravindrakumar (2023). Adaptive Intrusion Detection Systems Using Machine Learning in Cloud Environments. *ShodhKosh: Journal of Visual and Performing Arts, 4*(1), 1272–1278. doi: 10.29121/shodhkosh.v4.i1.2023.3358

1.1. SECURITY CHALLENGES IN CLOUD ENVIRONMENTS

Cloud infrastructures are extremely dynamic and multi-tenant, which makes them appealing targets for assaults because of their intrinsic characteristics. There are times when traditional security methods have difficulty keeping up with the constantly shifting threat landscape. Cloud computing systems are prone to a variety of problems, including data breaches, distributed denial-of-service (DDoS) assaults, and threats from within the organisation. In addition, the growing prevalence of encrypted communication and the sheer amount of data both contribute to the complexity of threat detection. These problems bring to light the necessity of security systems that are both intelligent and adaptable, and that are able to perform analysis and responses in real time.

1.2. ROLE OF MACHINE LEARNING IN INTRUSION DETECTION SYSTEMS

The application of machine learning has brought about a revolution in the field of intrusion detection by making it possible for systems to acquire knowledge from previous data and enhance their capability over time. ML-driven systems, in contrast to classic rule-based intrusion detection systems, are able to recognise attack pathways that were not previously identified and can adapt to new threat patterns. Anomaly detection, pattern identification, and predictive analytics are all areas in which intrusion detection systems (IDS) make extensive use of various learning techniques, including supervised learning, unsupervised learning, and reinforcement learning. These systems are able to discriminate between legal and harmful actions with a high degree of accuracy via the analysis of huge volumes of data. This results in a considerable reduction in the number of false positives and an overall improvement in security.

1.3. ADAPTIVE INTRUSION DETECTION IN THE CLOUD

The implementation of adaptive intrusion detection systems in cloud environments makes use of the capabilities of machine learning to handle the special security concerns that are associated with cloud computing. Monitoring network traffic and user behaviour in a continual manner, adaptive intrusion detection systems (IDS) constantly update their detection models in order to keep one step ahead of new threats. In order to provide a proactive approach to security, these systems are equipped with features such as self-learning algorithms and automatic reaction mechanisms, which enable them to minimise hazards in real time. A further advantage of adaptive intrusion detection systems is that they can grow with cloud workloads, which guarantees constant security across a wide variety of scenarios.

1.4. BENEFITS AND FUTURE PROSPECTS

The use of adaptive intrusion detection systems that make use of machine learning provides a multitude of advantages for cloud security. These enhanced threat detection capabilities guarantee that assaults are identified and mitigated in a timely manner, hence decreasing the potential harm that might occur. These systems lend themselves very well to dynamic cloud infrastructures due to their enhanced scalability and flexibility. In addition, the ongoing development of machine learning algorithms holds the potential of even better precision and effectiveness in the application of IDS in the subsequent years. The relevance of adaptive intrusion detection systems (IDS) will continue to expand as organisations become more reliant on cloud services. This will make adaptive IDS a fundamental component of the next generation of cybersecurity strategy. There has been a substantial breakthrough in cloud security brought about by the use of machine learning into intrusion detection systems. The machine learning-driven intrusion detection system (IDS) guarantees a secure foundation for digital innovation and growth by tackling the specific problems that cloud environments present and by delivering security that is both real-time and adaptable.

1.5. OBJECTIVE

1) To Evaluate and Contrast Various Machine Learning Models.
2) Use machine learning to find significant trends and anomalies to establish the most important features for cloud intrusion detection.

2. METHODOLOGY

The purpose of this research is to investigate and assess a number of machine learning (ML) techniques for the purpose of detecting intrusions in cloud settings. The selection of datasets, the engineering of features, the training of models, and the evaluation of their performance are all components of our technique, which follows a methodical approach. In order to determine which machine learning approaches are most successful in identifying cloud-based intrusions, the primary objective is to compare and contrast these strategies.

The Selection of Datasets: We make use of datasets that are open to the public and have been developed particularly for the purpose of intrusion detection in order to evaluate the effectiveness of the machine learning models. These datasets contain both normal and malicious traffic logs, and labels indicate whether an occurrence is benign or indicates an intrusion. These annotations are included in the datasets. The CICIDS 2017 dataset, which offers statistics on network traffic gathered from a cloud environment, is the major dataset that was utilised in this investigation. Several different attack scenarios, including distributed denial of service (DDoS) and other sophisticated persistent threats, are included in the dataset. This ensures that a diverse representation of incursions in cloud systems is achieved.

3. PREPROCESSING AND FEATURE ENGINEERING

The raw dataset is subjected to a number of preprocessing stages in order to guarantee the quality of the data and its relevance for the purpose of training machine learning models:

Data Cleaning: Through the use of imputation techniques, missing values, duplicates, and noisy data are either deleted or replaced.

Feature Selection: On the basis of domain knowledge and statistical approaches (such as correlation analysis and feature importance), a subset of characteristics is chosen to be used. Because of this, dimensionality is decreased, and model performance is enhanced.

Feature Scaling: It is possible to achieve optimal performance with some algorithms, such as Support Vector Machines (SVMs) and closest neighbors (KNN), by normalising or standardizing the features. This ensures that all of the input variables are on the same scale.

4. MACHINE LEARNING MODELS

The following machine learning algorithms are evaluated for their performance in intrusion detection within the cloud environment:

1) The Regression of Logistic A linear classifier that is designed to do binary classification jobs in an effective manner.
2) An example of a non-linear classifier that is simple to understand and offers several decision-making pathways is the decision tree framework.
3) The Forest of Randomness An ensemble model that makes use of many decision trees and then aggregates the predictions made by those trees in order to achieve greater precision.
4) (SVM) stands for support vector machines. A classifier that is aimed at locating the hyperplane that is most suitable for dividing data into these several categories.
5) It is a non-parametric approach that classifies an instance based on the majority class of its neighbours. K-Nearest Neighbours (KNN) is an acronym for this method.
6) Neural Networks (TNs) An technique to deep learning that makes use of multi-layer structures in order to understand intricate patterns concealed within the data

A total of seventy percent of the data is used to train each model, and the remaining thirty percent is used to test and evaluate the models.

5. MODEL EVALUATION METRICS

Several different assessment measures are utilised in order to ensure that the performance of the machine learning models is compared :

- Preciseness It is the percentage of cases that have been accurately categorises.
- To be precise in comparison to all positive forecasts, the percentage of predictions that turned out to be accurate.
- Don't forget the percentage of forecasts that turned out to be accurate out of the total number of real positives.
- An F1-Score Precision and recall are both achieved through the use of harmonic techniques, which provide a balance between the two.
- The region Under the Curve (AUC) is a statistic that provides an overall assessment of model performance by providing a summary of the trade-off between the true positive rate and the false positive rate.

Validation by Cross-Checking An application of 10-fold cross-validation is carried out during the training of the models in order to guarantee the robustness and generalizability of the models. In order to do this, the dataset is divided into ten subgroups, the model is trained on nine of these subsets, and the model is validated on the remaining subset. The procedure is carried out ten times, with each subset acting as the validation set once, and the average performance metrics are computed after each iteration.

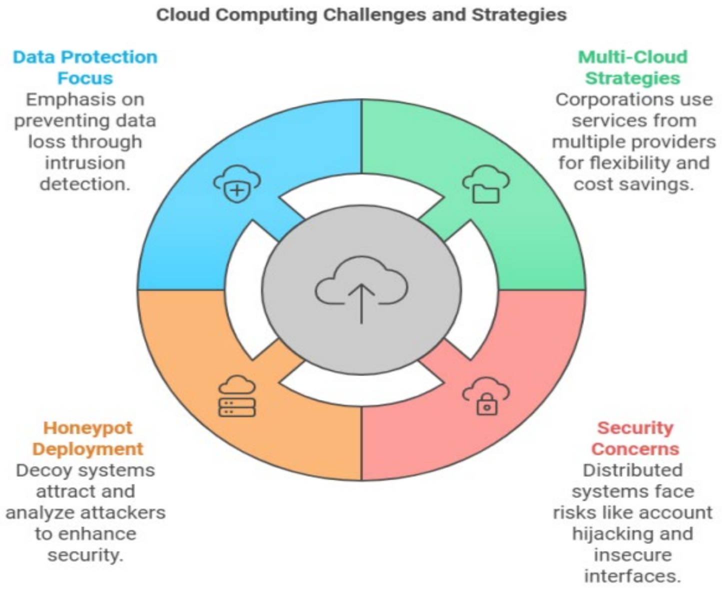
6. COMPARISON OF RESULTS

All models' performance characteristics are evaluated to find the best cloud intrusion detection model. The viability of any model for deployment in real-world cloud systems is also determined by computational efficiency, such as training and inference time.

Cloud computing is growing quickly in the commercial and governmental sectors, including enterprises, corporations, and governments worldwide. Cloud systems provide three main service models: IaaS, SaaS, and PaaS. Cloud companies provide several services based on these ideas. Due to service range, quality, pricing, and performance, global cloud service providers compete fiercely. Microsoft Azure, Google Cloud, AWS, Oracle, IBM, and others are market leaders. Three operating environments—private cloud, public cloud, and hybrid cloud—categorize the accessibility and utilisation of various service types. Cloud services were initially organised, but rising user base has led to more customised settings. Companies increasingly mix services from diverse suppliers based on pricing, quality, and performance. This strategy created the multi-cloud environment, where workloads are spread among infrastructures and computing resources. Multi-cloud techniques save money, increase disaster recovery, allow flexible corporate planning, and boost operational efficiency. They also provide issues including assuring data accessible across infrastructures, enforcing similar data rules across cloud providers, and sustaining data availability with a sustainable user base.

Multi-cloud deployments are complicated, and security is a major issue. A single security technique may not work across all units due to the system's distribution across cloud platforms and services. In this dispersed system, encryption methods that need continual ciphering and decoding may not work. The biggest cloud computing security vulnerabilities include account hijacking, service theft, unsecured interfaces, and shared APIs. Security breaches in multi-cloud environments might have serious repercussions. A successful service incursion might jeopardise underlying infrastructure. An IaaS infiltration might allow the attacker to view virtual machine monitors and manipulate IaaS-provided virtual machines. Since cloud computing is distributed and shared, establishing a security framework for anomaly detection and privacy management is difficult. Cloud providers' lack of openness prevents customised intrusion detection systems from engaging with the service administration layer, making virtual instance monitoring and security difficult. This is why most intrusion detection systems are developed on big networks, but implementing and pilot testing them on cloud platforms, especially in a multi-cloud setting, is difficult. A honeypot is an efficient intrusion detection and prevention tool that attracts and studies intruders. It is designed to seem like a legitimate system or server with convincing folders, files, and information to lure the attacker. The system uses firewalls and intrusion detection systems to log the attacker's activity for examination. Honeypots aim to deflect hackers off the real network, construct criminal profiles, find new vulnerabilities, and capture malware for study. A honeynet is formed by many honeypots.

Data security challenges have grown with digital networks and scattered contexts. Traditional networks prioritised data at rest security, which led to resilience approaches. The internet and cloud computing have changed the focus. Data protection, cryptography, and security are important, but intrusion detection—identifying abnormalities and unauthorised network access—is increasingly important nowadays. Many frameworks and methods have been developed for this essential study topic. The Common Intrusion Detection Framework (CIDF) defines functional intrusion detection modules. This section briefly describes CIDF and its operations.


Cloud Computing Challenges and Strategies
Data Protection Focus
Emphasis on preventing data loss through intrusion detection.
Multi-Cloud Strategies
Corporations use services from multiple providers for flexibility and cost savings.
Honeypot Deployment
Decoy systems attract and analyze attackers to enhance security.
Security Concerns
Distributed systems face risks like account hijacking and insecure interfaces.

7. COMPARATIVE ANALYSIS

Many tools and methods have been developed to solve multi-cloud security issues. Intrusion detection systems (IDS) are becoming increasingly flexible to fit multi-cloud designs. Snort (www.snort.org), SPADE, LAD, Prelude, and Stealth watch (now Cisco Secure Network Analytics and Breach Gate) have been updated to capture user behaviour, login patterns, and routine anomalies. Distributed sensors and agents monitor normal and aberrant network behaviour with these tools. IDS tools can also be divided into intrusion detection and prevention systems, integration tools, and service-specific tools.

The complexity of multi-cloud security concerns divides intrusion detection methods into three primary categories: statistics, knowledge, and machine learning. Each method has pros and cons.

Statistical-Based Models: They are split into univariate, multivariate, and time-series models. They create two stochastic datasets from network traffic activity. IP addresses, traffic rate, protocol data packets, and connection rates are in these databases. Comparing the dataset to a network statistical profile identifies intrusions. If the comparison score exceeds a threshold, an intrusion is identified.

Knowledge-Based Models: These models include FSM, UML, and expert systems. Knowledge-based IDS uses network data to determine essential traits and categories to create classification rules and parameters. These systems are manually trained and use rule bases to set intrusion detection levels. They decrease false positives during training, but new and unanticipated threats may not be covered after training.

Machine Learning-Based Techniques: This fast-growing field covers Bayesian Networks, Markov Models, Neural Networks, Fuzzy Logic, Genetic Algorithms, and Clustering for outlier discovery. Labelled data is needed to create pattern recognition and classification models, which is resource-intensive. Statistical models may be strengthened using machine learning techniques by integrating components from other categories, lowering processing costs.

Three popular cloud computing architectures for multi-cloud setups are HAIL, RACS, and ICStore. Each has pros and cons.

HAIL (High Availability and Integrity Layer): HAIL, presented by K.D. Bowers in 2009, manages file systems between cloud services and servers. It lets people access files across cloud platforms without protocol modifications. HAIL uses a proxy service to connect servers and cloud services for users. Cryptographic aggregation ensures file integrity even if portions of the multi-cloud system are hacked. HAIL can't manage file versioning or dynamic file systems, which is its biggest drawback.

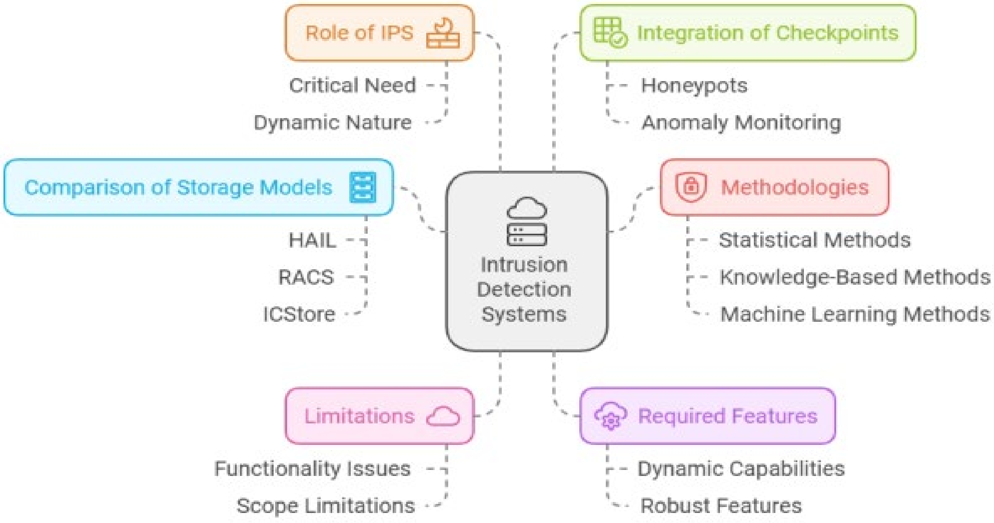
RACS (Redundant Array of Cloud Storage): This architecture handles multi-cloud storage to find consumers the most cost-effective and secure resources. RACS uses a distributed file management system across many cloud services and providers like RAID5. The main benefits of RACS are availability, replication, and efficiency across cloud platforms. HAIL's main drawback is the absence of file versioning, while RACS manages distributed storage better.

ICStore (InterCloud Storage): In 2010, Cachin et al. created ICStore to protect data's confidentiality, integrity, dependability, and consistency (CIRC) in multi-cloud environments. ICStore has stronger security and more exact specifications than HAIL and RACS. ICStore handles security issues better than HAIL and RACS via asynchronous, fault-tolerant client-driven storage protocols. HAIL employs symmetric cryptographic keys, which users must safeguard, whereas RACS and ICStore use RAID5 to handle distributed storage between servers and cloud services.

According to the findings of the comparison research, the present intrusion detection systems (IDS) have limitations in terms of both their functioning and their breadth. These technologies could be useful for deployments that just include a single cloud, but they do not possess the dynamic skills necessary to manage situations that involve several clouds. In order to effectively manage security in a multi-cloud structure, intrusion detection systems (IDS) need to contain more robust capabilities to meet the complexity of managing numerous clouds and services, as shown in Table 1.

Table 1 Comparing Different Storage Models

Model	Service	Feature	Summary
HAIL	Storage	Encryption Key	Strong security but lacks file versioning
RACS	Storage	RAID5	Strong distribution but low security
ICStore	Storage	CIRC	Strong distribution with reasonable security



Within the context of cloud computing settings, this study investigated a variety of factors pertaining to intrusion detection and security concerns. It is of the utmost need to take into consideration the ever-increasing complexity of incursions, which include anything from data-centric attacks to application-based vulnerabilities to fundamental cloud services. It is abundantly clear that traditional intrusion detection systems (IDS) are not as important as intrusion prevention systems (IPS) when it comes to multicloud settings because of the dynamic nature of these environments. In multi-cloud configurations, the statistical, knowledge-based, and machine learning techniques each provide useful characteristics for both detection and prevention. These methodologies also help in the prediction of attack patterns and analytics, which leads to more efficient management and strategy formulation.

8. CONCLUSION

the growing significance of strong intrusion detection and prevention systems in cloud computing's dynamic and complex environment, especially in multi-cloud settings. It is evident that more proactive intrusion prevention systems must be used in conjunction with traditional intrusion detection systems for successful threat mitigation as incursions target several levels, ranging from cloud services to data and application vulnerabilities. Cloud security may be improved by utilising statistical, knowledge-based, and machine learning approaches to identify and address risks as well as anticipate possible attack patterns, which enables more effective resource allocation and strategic planning. The robustness of cloud systems may be greatly increased by combining machine learning for clustering and pattern recognition with checkpoints for anomaly monitoring and honeypots. Adopting such cutting-edge security techniques across all service levels (SaaS, PaaS, and IaaS) will be essential for responding to new threats and maintaining a proactive and flexible security posture as multi-cloud systems continue to expand and diversify. In the end, this study establishes the foundation for developing intelligent and flexible cloud security tactics that will help the cloud ecosystem remain ahead of increasingly complex cyberthreats.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Ahmed, E., & Maher, G. (2024). Optimizing Supply Chain Logistics with Big Data and AI: Applications for Reducing Food Waste. Journal of Current Science and Research Review, 2(02), 29-39.

Gerges, M., & Elgalb, A. (2024). Comprehensive Comparative Analysis of Mobile Apps Development Approaches. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 6(1), 430-437.

Gerges, M., Elgalb, A., & Freek, A. (2024). Concealed Object Detection and Localization in Millimetre Wave Passengers' Scans. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(4), 372-382.

Zolotan, M., & Ross, A. (2016). Intrusion Detection Systems in Cloud Computing: A Survey. International Journal of Computer Applications, 143(10), 1-5.

S. Pal, S. Khatua, N. Chaki, and S. Sanyal, "A New Trusted and Collaborative Agent Based Approach for Ensuring Cloud Security", International Journal of Engineering 2012.

Zhe Li, Weiqing Sun, Lingfeng Wang, "A Neural Network Based Distributed Intrusion Detection System On Cloud Platform", IEEE 2012.

Changsheng Xiang , Zhou Yu , Xilong Qu, "Support Vector Machine Optimized by Improved Genetic Algorithm" Telkomnika Indonesian Journal of Electrical Engineering 2014

Preeti Mishra, Emmanuel S. Pilli , Vijay Varadharajan, Udaya Tupakula, "Nucleoids: A Security Architecture to Detect Intrusions at Network and Virtualization Layer in Cloud Environment", Conference on Advances in Computing, Communications and Informatics 2016.

Preeti Mishra, Emmanuel S. Pilli, Vijay Varadharajan,Udaya Tupakula, "Efficient Approaches for Intrusion Detection in Cloud Environment", International Conference on Computing, Communication and Automation (ICCCA2016).

Kleber, schulter, "Intrusion Detection for Grid and Cloud Computing", IEEE Journal: IT Professional, 19 July 2010.

Modi, C., Patel, D., Patel, H., Borisaniya, B., Patel, A. and Rajarajan, M. (2013). A survey of intrusion detection techniques in Cloud. Journal of Network and Computer Applications, 36(1), pp. 42-57. doi: 10.1016/j.jnca.2012.05.003.

Chirag Modi, Dhiren Patel, Bhavesh Borisanya, Avi Patel, and Muttukrishnan Rajarajan. A novel framework for intrusion detection in cloud. In Proceedings of the Fifth International Conference on Security of Information and Networks, pages 67– 74. ACM, 2012.

K.Deepa, and M.Chatterjee. "An adaptive distributed intrusion detection system for cloud computing framework." In Recent Trends in Computer Networks and Distributed Systems Security, pp 466-473, Springer, Berlin, Heidelberg, 2012.

Marwane Zekri, Said El Kafhali, Noureddine Aboutabit and Youssef Saadi. "DDoS attack detection using machine learning techniques in cloud computing environments." Conference: 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), 2017.

Amar Amouri, Vishwa T. Alaparthy and Salvatore D. Morgera. "A Machine Learning Based Intrusion Detection System for Mobile Internet of Things." Advanced Intrusion Detection & Mitigation Systems in Wireless Sensor Networks, Sensors 2020, 20(2), 461.