CLOUD SECURITY ANOMALY DETECTION USING AI: A REAL TIME ANOMALY DETECTION SYSTEM FOR CLOUD WORKLOADS

Shivani Patil[1*], Anisha P Rodrigues[2]

[1*]*Student, Department of CSE, NMAM Institute of Technology, Nitte*
Email I'd: shravanipatil28007@gmail.com
[2]*Associate Professor, Department of CSE, NMAM Institute of Technology, Nitte*

Abstract
Cloud computing has been integrated into the core of contemporary computing infrastructures, and it has allowed organizations to effectively support large and dynamic workloads. Nevertheless, cloud environments have become more complex and diverse, posing serious security issues. Conventional intrusion detection systems, which make use of preset rules or known attack signatures, do not reveal the unknown or developing threats. In this paper, the researcher will present Intelli Guard Cloud, an artificial intelligence-based real-time anomaly detection system to improve the security of the cloud environment. The system continuously observes the workloads in the cloud and multi-source telemetry, such as log records of the system, network traffic, API activity, and metrics of resource use. It uses a hybrid detection model based on the combination of an Autoencoder model, an Isolation Forest model and a Long Short-Term Memory (LSTM) model to identify statistical, structural, and temporal anomalies. The suggested framework is tested on simulated and real-life cloud workload data under simulated attack conditions. The experimental findings point to the fact that the system delivers better detection accuracy, lower-false positive and low-latency response in comparison with the traditional approaches. These results prove how AI-based solutions can be effective in providing adaptive and scalable anomaly detection to dynamic clouds.

Keywords- Cloud Security, Anomaly Detection, Machine Learning, Deep Learning, Real-Time Monitoring

Received: 04/03/2026 *Revised: 15/04/2026* *Acceptance: 22/04/2026* *Publication: 28/04/2026*

1. INTRODUCTION

Over the last few years, Amazon Web Services (AWS) and Google Cloud have emerged as cloud computing platforms that are becoming a part of contemporary organizational infrastructures. The types of workloads supported on these platforms are many and include virtual machines, containerized applications, microservices, and serverless architectures. Although cloud computing comes with scale and flexibility, it also comes with considerable security risks that need to be mitigated to secure and reliable operations.

Cloud environments generate huge volumes of heterogeneous data, including system logs, API calls, network logs, resource usage monitors, and more. Such data is extremely dynamic and constantly changing, and it is hard to establish some fixed rules on how to effectively detect threats. In the recent past, much has been done in regard to DDoS (Distributed Denial of Service) attacks in the framework of modern network systems since it implies that there must be efficiency in the detection systems [1]. The latest innovations in artificial intelligence and deep learning have allowed creating more adaptive and data-driven intrusion detection systems. The strategies use bulk volumes of data to improve on improving detection of all types of cyber threats including sophisticated and dynamic forms of attacks [2].

Anomaly based methods of intrusion detection, as well as conventional intrusion detection systems have been studied extensively in the literature. Nevertheless, a lot of these methods are constrained to deal with dynamic environments, high dimensionality data, and changing patterns of attacks especially when working with large scale distributed systems [3]. Moreover, standard rule-based and signature-based approaches are also restricted to identify the zero-day attacks and normally record high false positive rates thus inefficiency in the real world of the cloud [4].

In a bid to overcome such problems, AI-based anomaly detection systems have become a viable solution. Such systems based on the historical data learn the normal behavior of systems and detect abnormalities that could be the signs of possible threats. Autoencoders and Isolation Forests machine learning models are useful at identifying anomalous patterns, and long-term forecasts in cloud workload behavior can be identified with deep learning models, mostly Long Short-Term Memory (LSTM) networks.

An anomaly detection system operates on a continual basis to monitor all the data (logs, metrics, and network flows) produced by cloud monitoring services and uses this information to identify an anomaly or deviation in learned patterns. When abnormal activity is detected, alerts are sent to facilitate quick action to be taken. These systems are commonly implemented on scalable systems, such as microservices and container frameworks, to perform efficient processing of large streams of data with low latency.

2. RELATED WORK

The past anomaly detection methods used in the field of cybersecurity were mostly based on statistic methods, threshold methods as well as rule-based intrusion detection systems. Although such techniques are computationally fast and

effective at monitoring established patterns of an attack, these techniques do not have the capability to monitor unknown or changing threat in dynamic clouds. As cloud sprinkles grow more complex in nature, the current study has taken a direction towards machine learning approaches to detect anomalies in a more flexible and evidence-based manner [5].

It is on this basis that significant attention has been accorded to deep learning-based techniques since they can model complex and high-dimensional data. Thorough research by Ferrag et al. [6] on deep learning applications in cyber security intrusion detectors is an example that demonstrates effectiveness of deep learning to detect intricate patterns, and how deep learning is more effective in identifying intruders compared to traditional ones. However, these approaches have a reputation of posing problems of computational complexity and semantic understandability of models.

The Real-time Anomaly Detection System (RADS) suggested by Ahmed et al. [7] is a remarkable contribution to the field of real-time anomaly detection systems. RADS is a one-class classifier that uses time-series analysis to identify abnormal behavior at the virtual machine. It possesses better detection accuracy along with a lower false alarm rate; thus, there is an added significance of applying temporal analysis for monitoring cloud security.

In the newer approaches, such as CloudShield, the capability of real-time anomaly detection has been provided due to its architecture which is scalable and adaptable for dealing with the fast-moving stream of information [8]. This demonstrates the importance of designing efficient detection architectures in the context of real-time situations.

Apart from infrastructural methods of detection, newer studies have also considered user behavioral analysis and unusual access patterns as tools for insider threats detection. The techniques examine the activity of users after a certain time and identify the anomalies to the behavioral descriptions. Although these improvements are made, the current solutions continue to have a number of limitations such as high computational cost, low explainability, and difficulties in matching the heterogeneous and dynamic environment of the cloud.

3. MOTIVATION AND RESEARCH CHALLENGES

There has been a great change in the modern computing infrastructures owing to the swiftness of cloud computing that allows organizations to compete in terms of scalability, flexibility and cost efficiency. Data storage and application deployment as well as processing at scale are now commonly handled in cloud. Nevertheless, this popular usage of cloud services also leads to significant security issues. New threats to the security of clouds may lead to dire outcomes, such as data breaches, financial loss, disruption of services and reputation damages.

The first reason why this study was conducted is because of the shortcomings of conventional security systems that are used in clouds. Traditional techniques, like rule-based and signature-based intrusion detection system, are fixed and rely on a set of known attack patterns. As a result, they are ineffective in handling:

- rapidly evolving and dynamic cloud workloads,
- high-volume, high-velocity streaming data, and
- never seen or zero-day attacks.

Moreover, there is a surge of advanced and multi-level threats in the current cloud environments, such as insider attacks, cryptojacking, sideways movements, and sleek resources exploitation. Such attacks can be subtle and dynamic in nature and tend to go under the radar of systems that have fixed rules.

Despite the obvious strengths of such AI-driven solutions when it comes to addressing those gaps, there are some challenges associated with their use for monitoring cloud architecture, which should be considered.

3.1 Data Imbalance

First, the distribution in cloud monitoring tends to be rather skewed since the number of normal data points far exceeds that of anomalies. Such an unbalanced dataset can bias machine learning models toward common examples. As a result, a significant problem is how to develop models capable of accurately identifying minority anomalous events without raising false positives.

3.2 Concept Drift

Cloud workloads are dynamic in nature, their behavior patterns are changing over an extended period because of scaling, workload migration, changes in software and changing user requirements. The actual phenomenon is called concept drift, and this may diminish the performance of the model when the learning framework fails to conform to novel data distributions. This means detection systems which identify anomalies should include a continuous learning mechanism or a periodical updating of models to ensure that the methods are effective in the long term.

3.3 Real-Time Processing Requirements

Cloud resources produce enormous amounts of data at very fast speeds, and real-time or close to it processing is required to efficiently detect threats. Located in the latter half of the process, delay may cause attacks to rapidly escalate, especially in the situations of either DDoS or cryptojacking. Therefore, a low-latency anomaly detection with high accuracy is a mandatory issue in making the detection practically viable.

3.4 Privacy and Data Security Concerns

Cloud telemetry data could contain confidential data, such as logs of activity by users or system activities. Distributing and processing this type of information to distributed systems may create privacy and compliance issues. The security of data processing and the lack of exposure to sensitive data is thus a vital aspect in designing anomaly detection structures.

3.5 Research Objective

To address the challenges, the study will focus on creating an AI-based anomaly detection system that can:

- properly detecting known and unknown threats contained in cloud workloads,
- adjusting to changing patterns of data and workloads,
- and operating in real-time with a low detector latency.
- Presentation of a reliable and scalable solution that fits in the contemporary cloud settings.

The IntelliGuard Cloud system that has been proposed will solve these challenges by implementing a hybrid modeling scheme, which allows improving the performance of detection, high-resilience, and operational viability of the cloud architecture on large scales.

4. PROPOSED METHODOLOGY

IntelliGuard Cloud is a proposed system that is an anomaly detection and a real-time tool that allows securing cloud workloads against abnormalities and possible malicious actions. The model incorporates machine learning and deep learning to detect abnormal workload behaviours throughout changing cloud environments. Compared to the traditional rule-based security systems IntelliGuard Cloud trains behavioral patterns using multi-source cloud telemetry and conducts continuous anomaly evaluation with low-latency alerts.

The general steps of the proposed methodology are five key steps: data collection, data preprocessing, model training and anomaly detection, real-time monitoring, and automated response generating alerts. The architecture is suitable to enable scalable deployments in current cloud solution environments, such as virtual machines, containers, microservices, and API-driven services.

4.1 Data Collection

The initial phase of the framework implies gathering data related to operations and security of interest across various sources of clouds. Given the fact that cloud workloads produce non-homogeneous and large-scale telemetry data, the system bundles facts that are obtained through various channels of monitoring to obtain a holistic view of system behavior. The data collected are:

- system logs,
- network traffic records, And Traces of API activity.
- CPU usage, memory usage, disk usage and bandwidth usage.

Through cloud monitoring agent, log collectors and telemetry pipelines, these data sources are constantly obtained. Such a combination of sources allows the framework to provide information on the infrastructure level and application level behavioral traits and thus enhances the ability to detect anomalies at various attack surfaces.

4.2 Data Preprocessing

However, the collected cloud telemetry data could be noisy, unstructured, and varied in formats. Therefore, a specific preprocessing technique needs to be applied before the training and prediction processes. Such a process is crucial in improving the quality of data and making sure that the detection models will be compatible. The preprocessing process involves the following steps:

1. Noise removal: Many irrelevant, duplicated, and erroneous records are stripped away from the raw data stream.
2. Data normalization: Numbers are normalized so that the disparity in the magnitude would not create a bias.
3. Format standardization: Metrics, traffic logs, and other records are transformed into a structure that can be processed through machine learning algorithms.
4. Feature extraction: The features that are relevant to the workload behavior are obtained including request frequency, response delay, CPU spiking, memory anomalies, access frequency, login attempts, and traffic variation.

In sequential analysis, time-varying observations are grouped into time windows in such a way that temporal short term and long-term workload behaviour may be effectively captured. The latter is particularly significant to the LSTM-based constituent of the framework.

4.3 Model Training and Anomaly Detection

IntelliGuard Cloud uses a combination of three complementary models, Autoencoder, Isolation Forest, and Long Short-Term Memory (LSTM) to enhance its robustness and detection accuracy. The models represent two different aspects of abnormal behavior and their combination diminishes the shortcomings of using an individual detector.

4.3.1 Autoencoder-Based Detection

The autoencoder model learns the compact representations of normal workload behavior of clouds. It is trained on non-pathological or mostly non-pathological sets of data and tries to rebuild the input at the output node. Anomalous observations are identified in the inference with the aid of the reconstruction error. Having a high reconstruction error means that the pattern observed is so different as compared to the pattern that has been learned before.

Let $x$denote an input feature vector and $\hat{x}$its reconstructed representation. The reconstruction loss is computed as:

$$L_{AE} = \| x - \hat{x} \|^2$$

If L_{AE}exceeds a predefined threshold, the corresponding sample is flagged as potentially anomalous.

4.3.2 Isolation Forest-Based Detection

Isolation Forest is used as an unsupervised anomaly detector method to detect rare and isolated observations of the cloud telemetry. The model generates numerous random decision trees and pins out the anomalous points based on shorter paths than typical cases. Given that anomalies are statistically less and diverge, they are distinguished faster than normal cases.

This model is computationally effective and especially applicable in the high-dimensional cloud monitoring data, which is useful in the security analysis in real-time.

4.3.3 LSTM-Based Temporal Detection

Since workloads in clouds have strong temporal correlations, the LSTM network is included to address the sequential behavior over time. Time-windowed sequences of telemetry are processed by the LSTM, and normal temporal behavior in the workload evolution is learned by the LSTM. Abnormalities in the expected patterns are observed as an anomaly, e.g. sudden changes in the intensity of traffic, repetitive incidents of failed access or unusual trends of resource consumption.

The LSTM module specifically provides a good performance in identifying attacks that follow over time such as insider abuse, cryptojacking, lateral movement, and resource abuse timely.

4.4 Ensemble Anomaly Scoring

All three models provide an anomaly score of the observation or succession that is entered. Ensemble fusion mechanism is used to combine the individual scores to yield a better final decision. Where S ae, S i f, and S l stm are the scores of the anomaly delivered by the Autoencoder, Isolation Forest, and LSTM models, respectively. The anomaly score is obtained as:

$$S_{final} = w_1 S_{AE} + w_2 S_{IF} + w_3 S_{LSTM}$$

where w_1, w_2, and $w_3$represent the weights assigned to each model such that:

$$w_1 + w_2 + w_3 = 1$$

An observation is classified as anomalous when:

$$S_{final} > \tau$$

where $\tau$is the anomaly decision threshold determined during validation. The multi-ensemble approach proves to be effective in improving the stability of detection by combining reconstruction-based, isolation-based as well as temporal-sequence-based evidence of anomalies.

4.5 Real-Time Monitoring Framework

After being trained, the integrated model is then used in a real-time monitoring pipeline, which continuously examines the incoming cloud telemetry. The architecture is developed to handle streaming information with the least amount of latency thus allowing immediate security evaluation of running workloads.

The incoming events are sent to the preprocessing layer followed by sending the events to the three detection models in parallel to make inferences. The obtained anomaly scores are combined in real time and suspicious events are immediately identified. This architecture enables persistent observation of very dynamic cloud environments where workload patterns can vary at high speeds as autoscaling, orchestration or service migration occurs.

The given framework can be used with microservice-based and container-based cloud implementations and can be executed in a scaled manner over distributed infrastructure.

4.6 Alert Generation and Automated Response

A system anomaly will cause a response and alert system which will facilitate a prompt mitigation of the threat. The response layer will also be set up to be inter-operable with the existing cloud security and monitoring tools such as SIEM systems, log analysis tools, and orchestration services. When anomalies are identified the framework can perform the following actions:

- create security alerts to the administrators,

- send out alerts to monitoring and incident response systems,
- isolate suspicious workloads,
- block IP addresses or sessions that are malicious, and
- launch predetermined automated containment measures.

The latter response ability can not only enable the proposed system to detect suspicious behavior but also assist in proactive security enforcement in the current cloud environment.

4.7 Methodological Significance

Several advantages of using the proposed method for anomaly detection in cloud security are evident. For one, it unifies various types of telemetry data, providing better visibility into the behaviors of workloads. In addition, the approach combines the principles of statistical isolation, reconstruction learning, and time-series models that enhance the detection capabilities of various threats. Third, the real-time architecture provides the facility of quick alerting and maintainability of operations in huge clouded systems. Finally, the system is streamlined to suit the dynamic and evolving of the cloud workloads which is required to identify new attacks.

5. RESULTS

5.1 Experimental Setup and Datasets

To test the proposed IntelliGuard Cloud framework, a mixture of synthetic cloud workload data, actual cloud monitoring data, and simulated attack cases were used to make sure the cloud framework is robust in a variety of operational situations. The synthetic data was created to provide a simulation of dynamic clouds with controlled injected anomalies whereas the real-world data were telemetry measurement of cloud monitoring systems, which comprises system logs, network traffic, API interactions, and resource utilization metrics.

In order to provide life-like threat conditions, the several attack scenarios were also provided, such as the Distributed Denial of Service (DDoS), cryptojacking, unauthorised access attempts, and pattern of abnormal resource consumption. To capture sudden and progressive changes in workload behavior, these scenarios were created.

Experiments were all done in a setting that closely matched actual cloud infrastructure, with workload changes based on time and streaming data pipelines. The pipeline of evaluation was used to make sure that the trained models could be tested on unseen data to evaluate generalization potential.

5.2 Quantitative Performance Analysis

The hybrid anomaly detection approach proposed demonstrated high performance levels based on different evaluation metrics, including detection pain and false positives and reaction time. The integration of Autoencoder, Isolation Forest and LSTM model helped the approach not only to detect statistical outliers but also to recognize changes over time in cloud workload anomalies.

It always exhibited high accuracy when detecting anomalies, as it was capable of distinguishing between normal and abnormal activity regardless of data distribution. Also, there were no many false positives, and thus the proposed solution would be capable of avoiding unnecessary alarms, which is especially important in a cloud security environment.

The ability to detect anomalies in a matter of milliseconds due to low-latency characteristics is an advantage of the proposed approach as it allows responding to the changing nature of threats such as DDoS attacks and cryptojacking.

The framework successfully detected a wide range of attack types, including:

- sudden spikes in network traffic associated with DDoS attacks,
- abnormal CPU and memory utilization patterns indicative of cryptojacking,
- irregular access patterns linked to unauthorized login attempts, and
- anomalous API usage behaviors.

These results demonstrate the effectiveness of combining multiple learning paradigms for comprehensive anomaly detection in dynamic cloud environments.

5.3 Ablation and Component Analysis

An ablation study was used to assess the input of each model in the hybrid framework by examining the performance of each component and their combinations. These findings show that all the models play different roles in total detection ability.

The Isolation Forest part was effective in detecting statistical outliers in data with a large number of dimensions, especially when there was an unexpected anomaly. The Autoencoder showed a high level of learning discrimination of subtle deviations of the learned normal behavior by differences between reconstruction errors. In the meantime, the LSTM model was far more supportive of detecting intriguing temporal anomalies, particularly the ones, which develop overtime.

On its own, each model was found to have shortcomings regarding the ability to capture some kind of anomaly. Nevertheless, the ensemble mode of combined detection led to an impressive increase in detectability and consistency. Integration of anomaly scores minimized biases based on models and increased the overall reliability of the systems. In addition, the system kept consistent performance at the time when one of its components showed poorer efficiency, illustrating the power of the ensemble design.

5.4 Comparison with State-of-the-Art Methods

The proposed IntelliGuard Cloud architecture was compared in concept to the current methods of anomaly detection, such as traditional rule-based schemes, single-model machine learning-based schemes and deep learning-based anomaly detec.

The proposed system exhibited better threat detection performance with previously unknown and changing threats than rule-based and signature-based systems due to its data-based learning algorithm. Compared to single-model methods, the offered hybrid framework delivered a better performance due to considering various aspects of anomaly behavior, such as the statistical, structural, and temporal manifestations of the anomalies.

The new system also reflected superior response time and adaptation where real-time detection is involved because decisions need to be made quickly. Unlike traditional systems, which take a long time to detect and have high levels of false alarm, the new model reflected a good compromise between the two. In other words, there is clearly a measurable advantage of utilizing heterogeneous models compared to the present approach.

5.5 Scalability and Deployment Analysis

A conceptual comparison of the proposed IntelliGuard Cloud solution was made against the existing approaches to anomaly detection, including rule-based, single-model ML-based, and deep learning-based anomaly detectors.

The proposed solution showed superior threat detection capability with respect to novel and evolving threats due to the fact that it relied on a data-driven approach. When compared to single-model based approaches, the proposed framework provided enhanced performance because it considered multiple dimensions of the anomaly behavior, namely the statistical, structural, and temporal characteristics.

Moreover, the solution demonstrated better response times and adaptability, which is particularly relevant for applications with real-time detection capabilities. While classical solutions have notable latency in terms of detection and a high rate of false positives, the proposed model achieved a better compromise between the two aspects. Overall, the results obtained indicate that there is quantifiable value to heterogenous models in comparison to existing approaches.

5.6 Limitations and Trade-offs

While it is advantageous, there are some limitations to the system that have to be taken into account. First of all, combining multiple models will increase computational requirements, especially while training. Such a demand for computing power could pose issues in large-scale systems. Secondly, availability of suitable data will play a critical role in how well the system will perform. The effectiveness of the machine learning models used might suffer under circumstances where the data is either poorly labeled or of low quality. Thirdly, the accuracy of the detection process will be impacted negatively if there are noisy data present, especially in reconstruction-based models such as autoencoders. In addition, selection of model weights and anomaly threshold could impact the sensitivity of the system. Finally, while the proposed system has excellent detection capabilities, it lacks interpretability, especially with deep learning models like LSTM.

5.7 Future Research Directions

Future improvements might be considered regarding the efficiency, understandability, and adaptability of the suggested architecture. Optimization of the processing speed, for instance, by reducing the model's complexity and allowing the inference of more sophisticated models in real-time is one of the ways forward. One of the most important aspects to consider is how to make the model more understandable in the case of its practical application to security, where the ability to comprehend the cause of the anomaly is required.

Plugging encrypted information or privacy-sensitive data without breaching security is still a challenge. Federated learning or privacy-preserving anomaly detection are only some of the approaches that can be considered to help with this problem.

Lastly, the use of sophisticated architectures, such as graph neural networks (GNNs), will also be used to enhance the capacity of the system to learn how to classify complex relationships and dependencies within the cloud environments, especially in how to detect coordinated or distributed attacks.

6. Conclusion

This paper introduced IntelliGuard Cloud, a framework of artificial intelligence (AI)-supported real-time detection of anomalies in cloud workloads. The suggested system combines various machine learning and deep learning algorithms, such as Autoencoder, Isolation Forest, and LSTM, to reflect statistical, structural, and time-related trends of cloud behavior. Such a combination makes it more robust in detecting and allows detecting familiar and unfamiliar anomalies.

The experimental assessment indicates that the suggested framework can successfully achieve the reliable detection of anomalies with low false positive rates and the low detection latency, which can be suitable with the dynamic cloud setup. With constant monitoring of the multiple source cloud telemetry, the system effectively identifies critical threats mentioned as, DDoS attacks, cryptojacking and unauthorized access activities. Besides enhancing the accuracy of detection, the system facilitates the real-time response measures, and as a result, it allows proactive mitigation of

security threats before it intensifies. The scalable architecture also guarantees its use in clouds infrastructures of large scale and distributed clouds. Although effective, the system has computational overhead and needs high-quality data to be able to perform optimally. Future directions include the development of model efficiency, explainability, and the use of more advanced methods, including the explainable AI and graph-based learning. Overall, the offered solution emphasizes the possibilities of AI-powered solutions to enhance cloud security and provide intelligent threat recognition.

REFERENCE

1. Alarifi and A. Tolba, "Deep Learning for DDoS Detection in Software Defined Networking,".
2. S. S. Levy, M. A. Khan, and M. A. Ferrag, "Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparison," Journal of Network and Computer Applications, vol. 206, p. 103442, 2022. [Online].
3. M. Ahmed, A. N. Mahmood, and J. Hu, "A Survey of Network Anomaly Detection Techniques," IEEE Transactions on Network and Service Management, DOI: 10.1109/TNSM.2024.11011269, 2024.
4. P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, and E. Vazquez, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," Journal of Cluster Computing, vol. 27, pp. 1-25, 2023. [Online].
5. S. K. Sahoo, P. K. Das, and S. K. Rath, "AI-Based Anomaly Detection for Cloud Security Using Machine Learning Techniques," Journal of Computer Technology & Applications, vol. 9, no. 1, pp. 45-58, 2024. [Online].
6. M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep Learning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study," Journal of Cloud Computing, vol. 12, p. 91, 2023. [Online].
7. M. Ahmed, A. N. Mahmood, and J. Hu, "Real-time Anomaly Detection System (RADS) for Cloud Data Centres," arXiv:1811.04481, 2018.
8. J. S. Baek, K. H. Kwak, B. G. Lee, and S. H. Kim, "CloudShield: Real-time Anomaly Detection in the Cloud," NSF Public Access Repository, DOI: 10.1109/Cloud.2024.00000, 2024. [Online].