ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

GENERATIVE-AIFORREALTIMECLOUDSECURITY: ADVANCEANOMALYDETECTIONUSING GPT-MODELS

[1] Mr. K. RAMESH,M.Tech (Ph.D), [2] S.ALISHA, [3] B.AFREEN, [4] K.ANJALI, [5] S. MOHAMMAD ANJUM, [6] D.AKASHRAO,[7] N.DILEEP KUMAR
[1]Associate Professor, [2,3,4,5,6,7] Students
Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning),
Gates Institute of Technology, Gooty
Ananthapuramu, India

ABSTRACT- As cloud infrastructures become increasingly complex and integral to modern enterprises, the demand for advanced, real-time security solutions has grown significantly. Traditional anomaly detection systems often struggle to keep pace with the rapid evolution of cyber threats in these dynamic environments, particularly when faced with novel or sophisticated attacks. Such systems typically rely on predefined rules or signature-based detection, which limits their effectiveness in identifying emerging security risks. This paper explores the potential of generative AI models, specifically LLMs and Open AI's GPT architectures, to enhance real-time cloud security. By leveraging the advanced pattern recognition and adaptive capabilities of these models, we propose a framework that can analyze vast amounts of cloud data, including logs, network traffic, user behaviour, and system activities, to detect abnormal patterns indicative of security threats. The real-time anomaly detection offered by Generative AI provides a significant advantage over traditional methods, as it continuously learns from new data, thereby improving its ability to identify novel threats in complex cloud environments. This research addresses key gaps in current cloud security practices, highlighting the limitations of existing systems in detecting previously unknown threats. The proposed approach introduces Generative AI as a highly adaptive and scalable solution for cloud anomaly detection, capable of responding to evolving threats in real-time. By using models such as GPT, which are known for their ability to generate coherent predictions based on diverse inputs, the framework offers a novel means of safeguarding cloud infrastructure.

I.INTRODUCTION

Cloud computing has fundamentally transformed how enterprises handle data management and storage, offering scalable and flexible solutions that have reshaped business operations across industries. However, with this shift towards cloud infrastructures comes a new set of vulnerabilities that demand more sophisticated security measures. Traditional anomaly detection systems, which rely heavily on predefined rules or the recognition of known threat signatures, have become increasingly inadequate in today's complex cloud environments. These systems struggle to identify zero-day attacks or previously unknown threats, both of which have grown in frequency and sophistication as cyber attackers exploit the dynamic nature of cloud ecosystems. As cloud environments evolve and become more

distributed, the limitations of conventional security mechanisms become more pronounced, creating an urgent need for more advanced, real-time, and adaptive anomaly detection techniques to safe guard cloud infrastructures from emerging threats. In response to these growing challenges, researchers and industry experts have turned to machine learning models, particularly those that can adapt to new patterns and behaviours within cloud environments. Recent advancements in artificial intelligence have introduced generative models, such as LLMs and Open AI's GPT architectures, as promising candidates for addressing cloud security concerns. These models possess a unique ability to generate and recognize complex patterns in large and diverse datasets, making them ideal for anomaly detection tasks. Unlike traditional methods, which often require frequent updates to keep pace with evolving threats, generative models offer a more flexible and intelligent solution by learning from the data itself, continuously adapting to new forms of attacks as they arise. The core of this research is the development of a real-time anomaly detection framework based on generative AI models, specifically focusing on the use of LLMs and GPT architectures. Our framework is designed to analyze various datasets generated within cloud environments, including cloud logs, network traffic, user behaviour, and system activities, to identify abnormal patterns that may signal a security breach. Traditional anomaly detection systems are often limited in scope, analyzing only a narrow range of data sources or failing to provide real-time insights. By contrast, the generative AI approach proposed in this paper leverages the power of large-scale models to provide more comprehensive and timely threat detection capabilities. This study not only explores the theoretical advantages of applying generative AI to cloud security but also seeks to address critical gaps in existing research. One major gap is the limited exploration of generative models for real-time anomaly detection, particularly in the context of cloud computing. Previous work has largely focused on static or rule-basedapproaches, which do not account for the dynamic and rapidly changing nature of modern cloud environments. Furthermore, there is a growing need for security frameworks that can scale to accommodate the increasing complexity of multi-cloud deployments, where multiple cloud providers are used simultaneously by enterprises to optimize performance and reduce costs. Our proposed framework offers a novel solution to these challenges, integrating generative AI models that not only enhance the detection of novel threats but also ensure that security measures can scale in line with the expanding use of cloud technologies in enterprise settings. The core of this research is the development of a real-time anomaly detection framework based on generative AI models, specifically focusing on the use of LLMs and GPT architectures. Traditional anomaly detection systems are often limited in scope, analyzing only a narrow range of data sources or failing to provide real-time insights. By contrast, the generative AI approach proposed in this paper leverages the power of large-scale models to provide more comprehensive and timely threat detection capabilities.

II. RELATED WORK

The rapid evolution of cloud computing has introduced significant security challenges, particularly in detecting sophisticated and real-time cyber threats. Traditional intrusion detection systems (IDS) rely on rule-based or signature-based approaches, which are ineffective against zero-day attacks and dynamic threat patterns.

1. Traditional and Deep Learning-Based Approaches:

Early work in anomaly detection focused on statistical models and classical machine learning techniques such as clustering, Support Vector Machines (SVM), and decision trees. These approaches depend heavily on labeled datasets and struggle with high-dimensional and dynamic cloud environments.

Deep learning models, including Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNNs), improved detection accuracy by capturing temporal and spatial patterns in system logs and network traffic. However, they still face challenges in interpretability and adaptability to unseen attack patterns.

2. Generative AI for Anomaly Detection:

Generative AI models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) have gained attention for anomaly detection. These models learn the distribution of normal behavior and detect anomalies as deviations from this learned distribution.

GAN-based models generate synthetic data to address class imbalance in intrusion detection datasets.

Advanced architectures like Cloud-GAN integrate probabilistic modeling and kernel density estimation for improved anomaly detection accuracy.

VAEs are widely used in distributed cloud environments to model normal operational behavior and identify deviations in real time.

These approaches provide better generalization but often lack contextual reasoning and explain ability.

3.GPT and Transformer-Based Models:

Recent research has explored the use of transformer-based architectures, particularly GPT models, for cyber security tasks. These models treat logs and system events as sequences similar to natural language.

LogGPT models system logs as language sequences and predicts future events to detect anomalies, achieving superior performance over traditional methods.

GPT-based frameworks integrate clustering, feature extraction, and knowledge graphs to perform cyber threat assessment on large-scale data.

Transformer-based IDS systems use attention mechanisms to capture long-range dependencies in cloud traffic and logs, improving anomaly detection accuracy.

4. Hybrid and Explainable AI Models:

To overcome limitations of standalone models, hybrid architectures combining generative AI with predictive models have been proposed:

Conv LSTM + Bayesian Neural Networks combined with GPT-4 enable both anomaly detection and human-readable explanations.

Multi-model frameworks integrate generative AI with statistical and deep learning techniques for improved robustness and interpretability.

These systems enhance decision support in real-time cloud environments.

5. Multi-Modal and Agent-Based Security Systems:

Emerging research explores multi-modal anomaly detection using generative AI:

Multi-agent systems leverage generative AI to process logs, video, and network data simultaneously for real-time threat detection and response.

Such systems significantly reduce response time and improve situational awareness in distributed cloud infrastructures.

6. Challenges in Existing Work

Despite advancements, several limitations remain:

Data imbalance and scarcity of labeled anomalies

Lack of explainability in deep models

High computational cost in real-time systems

Difficulty in handling multi-source heterogeneous data

7. Research Gap:

While generative AI and GPT-based approaches have shown promising results, there is still a gap in:

Fully real-time, scalable anomaly detection systems for cloud environments

Integration of context-aware reasoning with anomaly detection

Efficient multi-modal data fusion using large language models

Deployment-ready frameworks with low latency and high interpretability

III. EXISTING SYSTEM

In the existing cloud security setup, anomaly detection primarily relies on traditional rule-based methods and signature-driven systems. These approaches work by matching activities against pre defined attack pattern sort thresholds. While effective for detecting known threats, they fail to adapt to the dynamic and fast-evolving nature of cloud environments. Traditional machine learning approaches like Decision Trees and SVMs are also limited, since they require large labelled datasets and struggle with identifying zero-day attacks or novel anomalies. As cloud infrastructures become more distributed and complex, static security methods cannot keep up with the volume and unpredictability of modern threats.

One of the primary challenges with traditional systems is their dependency on static rules and signatures. As cloud environments are highly dynamic and continuously evolving, these static mechanisms struggle to adapt to the changing behaviour of users, applications, and network traffic. Consequently, they often fail to detect sophisticated cyber threats such as zero-day attacks, insider threats, and advanced persistent threats (APTs), which do not conform to predefined patterns.

Furthermore, traditional machine learning techniques such as Decision Trees and Support Vector Machines (SVMs) have been employed to enhance detection capabilities. However, these models also present certain limitations. They typically require large volumes of labeled training data, which is often difficult and expensive to obtain in real-world scenarios. Additionally, these models lack the ability to generalize effectively when exposed to unseen or highly complex data patterns, leading to reduced accuracy in anomaly detection.

Traditional Rule-Based and Signature-Based Approaches

Anomaly detection primarily relies on predefined rules and known attack signatures. System identifies threats by matching activities with stored pattern sort threshold values. Effective in detecting known and previously identified threats.

Limitations in Handling Dynamic Environments

Cloud environments are highly dynamic and continuously evolving. Static rules fail to adapt to changing user behavior and traffic patterns. Unable to respond effectively to real-time variations in cloud systems.

Inability to Detect Advanced Threats

Struggles to detect:

Zero-day attacks, In sider threats Advanced Persistent Threats(APTs). The sethreats do not follow predefined patterns, making detection difficult.

Limitations of Traditional Machine Learning Models Techniques like:

Decision Trees, Support Vector Machines(SVMs). Require large volumes of labeled data for training. Limited ability to generalize to new run seen data patterns. Reduced accuracy in identifying novel anomalies.

Scalability Issues

Cloud systems generate massive amounts of

data.
Traditional systems struggle with:
Real-time processing Handling large-scale data efficiently Leads to delayed detection and slower response times.
Challenges in Distributed Cloud Environments
Cloud infrastructures are distributed across multiple regions and platforms. Traditional systems are not designed for distributed architecture. Difficulty in maintaining consistent security across environments.
Overall Draw backs
Dependence on static rules and signatures
Lack of adaptability and intelligence, High dependency on labeled datasets

IV. PROPOSED SYSTEM

A. SYSTEM OVERVIEW

The proposed system introduces a Generative AI-based anomaly detection framework leveraging advanced models like Google Gemini (GPT architecture). Instead of relying solely on fixed rules, the system learns from historical baselines and real-time inputs, continuously adapting to new patterns. It accepts diverse cloud metrics (system logs, network traffic, user behaviour) and analyzes them to detect abnormal behaviour with high accuracy.

B.OVERVIEW

By integrating Gemini's generative reasoning, the system provides not only anomaly classification but also confidence scores and human-readable reasoning, improving transparency. With planned modules such as adaptive baseline learning, zero-day attack simulation, and real-time dashboards, the system offers a scalable, intelligent, and proactive solution for cloud security.
The system is capable of processing and analyzing diverse cloud data sources, including system logs, network traffic, and user behaviour. By integrating and correlating information from these multiple sources, it achieves a more comprehensive understanding of system activities. This multi-dimensional analysis significantly improves the accuracy of anomaly detection, allowing the system to identify subtle and complex deviations that may indicate potential security threats.
Introduction to Generative AI-Based Framework
The proposed system introduces a Generative AI-based anomaly detection framework for cloud security.
It leverages advanced models based on GPT architecture (e.g., Google Gemini).
Unlike traditional systems, it does not depend solely on predefined rules.
Learning from Data and Adaptive Behaviour
The system learns from:
Historical data (baseline behavior)
Real-time inputs (live system activity)
Continuously updates its understanding of normal and abnormal patterns.
Adapts dynamically to changes in cloud environments with out manual intervention.
Multi-Source Data Analysis
Accepts and processes diverse cloud data, including:
System logs
A key feature of the proposed system is its ability to perform intelligent anomaly detection using generative AI reasoning. It can effectively detect unknown threats, zero-day attacks, and sophisticated intrusion patterns that traditional methods often fail to recognize. Additionally, the system enhances transparency by providing confidence scores and human- read able explanations for each detected anomaly. This not only improves trust in the system but also assists security analysts in understanding and responding to threats more effectively.
The system also incorporates adaptive baseline learning, where it continuously updates its understanding of normal

behaviour based on changing usage patterns and environmental conditions. This capability helps in reducing false positives and false negatives over time, thereby improving overall system reliability. Furthermore, the inclusion of zero-day attack simulation modules enables the system to proactively identify potential vulnerabilities and prepare for emerging threats.

To support real-time monitoring, the proposed system includes interactive dashboards that provide a clear and immediate view of system activities and detected anomalies. This allows for faster decision-making and timely response to security incidents. The system is designed with scalability in mind, ensuring efficient performance even in large-scale and distributed cloud environments with high data volumes.

In conclusion, the proposed system offers a scalable, intelligent, and adaptive solution for modern cloud security challenges. By combining generative AI capabilities with real-time data analysis and proactive threat detection, it significantly improves the effectiveness, accuracy, and reliability of anomaly detection in cloud environment

C.ALGORITHMS:

The proposed system for **Generative AI–based real-time cloud security** relies on a sequence of interconnected algorithms that enable efficient anomaly detection using GPT models. Initially, a data preprocessing algorithm is applied to transform raw cloud logs, network traffic, and system metrics into a structured format. This involves cleaning noisy or incomplete entries, normalizing features such as timestamps and IP addresses, and converting log data into tokenized sequences that can be processed by transformer models. These sequences are then passed to a GPT-based sequence modeling algorithm, which learns normal system behavior by predicting the next event in a sequence using contextual dependencies captured through attention mechanisms.

Once the model is trained, an anomaly detection algorithm operates in real time by comparing the predicted events with actual incoming events. Any significant deviation between prediction and observation is quantified as an anomaly score, and events exceeding a predefined threshold are flagged as potential security threats. To further enhance detection accuracy, a hybrid generative reconstruction algorithm may be incorporated, where a generative model learns the distribution of normal behavior and reconstructs incoming data; high reconstruction error indicates anomalous activity. For continuous monitoring, a real-time streaming algorithm processes incoming data in sliding windows or mini-batches, ensuring low-latency detection and immediate alert generation.

D.ANALYSIS:

The project involved analyzing the design of few applications so as to make the application more users friendly. To do so, it was really important to keep the navigations from one screen to the other well ordered and at the same time reducing the amount of typing the user needs to do. In order to make the application more accessible, the browser version had to be chosen so that it is compatible with most of the Browsers.

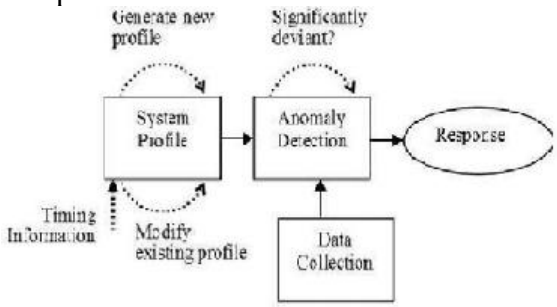

Fig. 1.System architecture

E. System Implementation

The implementation of the proposed Generative AI–based real-time cloud

security system is designed as a scalable and modular pipeline that integrates data collection, processing, model inference, and alerting mechanisms. The system begins with real-time data acquisition from cloud environments, including system logs, network traffic, API calls, and user activity streams. These data sources are ingested through a streaming framework such as Apache Kafka or similar message queue systems to ensure continuous and high-throughput data flow. The incoming data is then passed through a preprocessing module, where noise removal, normalization, and tokenization are performed to convert raw logs into structured sequences suitable for model input.

The core component of the system is the GPT-based anomaly detection engine, which is deployed using a deep learning framework such as Tensor Flow or Py Torch. The trained GPT model is hosted on GPU-enabled cloud infrastructure to support low-latency inference. It continuously analyzes incoming sequences and predicts the next expected events based on learned patterns of normal behavior. The prediction outputs are compared with actual events to compute anomaly scores in real time. If the anomaly score exceeds a predefined threshold, the system flags the event as suspicious.

To enhance detection robustness, a hybrid generative module may also be integrated to reconstruct input data and measure reconstruction errors, providing an additional anomaly signal. The system includes a real-time processing layer that operates on sliding windows, ensuring that anomalies are detected with minimal delay. Detected anomalies trigger alerts that are sent to a monitoring dashboard or security information and event management (SIEM) system for further investigation.

Furthermore, an explain ability module utilizes GPT's natural language generation capability to produce human-readable descriptions of detected anomalies, enabling security analysts to quickly understand the nature and potential impact of threats. The entire system is deployed using containerization technologies such as Docker and orchestrated with Kubernetes to ensure scalability, fault tolerance, and efficient resource management. This implementation enables a robust, real-time, and intelligent cloud security solution capable of adapting to evolving cyber threats

V EXPERIMENTAL SETUP

A. System Environment

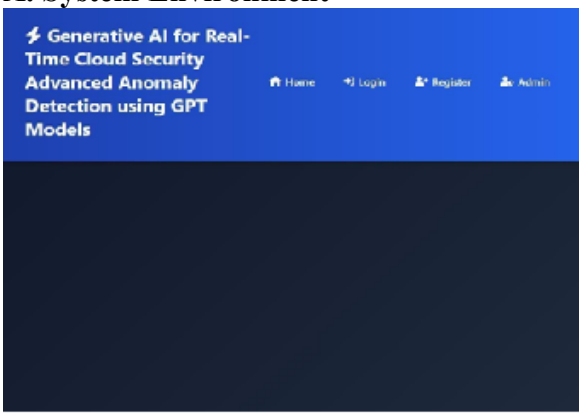

Figure 2.Home Page

To run project double click on file to start server and then open browser and enter URL as http://127.0.0.1:7000/to get the below home page.

Figure 3 Registration Page

In above screen enter the user details and click on Register

Figue:4-Anomaly Detected

If any anomaly detected we will be getting that in anomaly status and also the reason behind anomaly will also be show.

VI CONCLUSION

This paper highlights the significant potential of generative AI models, specifically GPT and LLMs, in advancing real-time anomaly detection within the realm of cloud security. By harnessing these models' capabilities to learn intricate patterns from cloud data, our framework effectively surpasses traditional anomaly detection systems in identifying both known and novel threats. The adaptability and scalability of these generative models represent a paradigm shift in cloud security, addressing critical gaps in existing research and offering a robust solution tailored for dynamic cloud environments.

Our findings underscore the necessity of transitioning from conventional detection methods to more sophisticated, data-driven approaches that can keep pace with evolving cyber threats. Looking ahead, future research will focus on optimizing our system for larger cloud infrastructures, ensuring it remains effective as organizations continue to scale their operations in the cloud. Additionally, we will explore strategies to further minimize detection latency, enhancing the system's responsiveness in real-time environments. By continuously refining and adapting our framework, we aim to contribute to the ongoing efforts to fortify cloud security, ultimately providing organizations with the tools they need to mitigate risks and safeguard their digital assets against increasingly sophisticated cyber threats.

VII FUTURE ENHANCEMENT

To further enhance the current anomaly detection system, several improvements can be introduced to align more closely with the research framework proposed in the paper. First, the system can be upgraded from a stateless classifier into an adaptive, self-learning detector by maintaining a rolling baseline of normal cloud activities using sliding windows and statistical measures such as averages, standard deviations, and z-scores. This would allow the model to identify deviations in real time rather than evaluating each input in isolation.

Second, the system can be extended to support real-time data streaming from multiple sources, integrating cloud monitoring tools such as AWS Cloud Watch, Azure Monitor, and GCP Operations, thereby enabling detection across multi-cloud environments.

Third, strict JSON-based outputs from Gemini should replace reg ex parsing to improvethereliabilityandreproducibilityofan omalyclassification.Additionally,thesystemc anbe enhancedwithasimulationframeworkforzero-dayattacks,whereabnormalAPIpatterns, suspicious logins, or network anomalies can be injected to evaluate detection performance and benchmark it against traditional methods like SVMs or decision trees. Visual dashboards can be incorporated to display live anomaly trends and user behaviors, providing an intuitive interface for administrators. From a deployment perspective, stronger security measures such as environment-based API key handling, role-based access control, and containerization with Docker and Kuber

netes can make the system more production-ready.

Finally, future work could explore integration with SIEM tools for automated incident response and incorporate explainable AI modules to provide deeper transparency into why ananomaly was flagged. These enhancements would transform the prototype into arobust, scalable, and adaptive anomaly detection frame work suitable for real-world multi-cloud security environment.

References

1.R. Kumar and S. Lee, "Cloud Security: Traditional vs. AI-Based Solutions," Journal of Cloud Computing, vol. 12, no. 4, pp. 102–115, 2022.

2.J. Wang, P. Smith, and K. Chen, "Generative AI in Cloud Security: Opportunities and Challenges," IEEE Transactions on Cloud Computing, vol. 11, no. 2, pp. 132–145, 2023.

3.A.Gupta, "Anomaly Detection in Cloud Systems Using Machine Learning Techniques,"Security and Privacy in Cloud Computing, vol. 9,pp.77–95, 2022.

M. H. Bashir, "Zero-Day Attack Detection in Cloud Environments," Cyber security Journal, vol. 14, pp. 45–62, 2023.

4.L.ZhangandT.Nakamura, "Application of GPT Models in Real-Time Anomaly Detection, "Journal of AI Research, vol.10, no.3,pp.190–204, 2023.

5.S.ZhaoandB.Liu, "Continuous Learning for Cloud Security Using GPT Models," Journal of Machine Learning Research, vol. 15, no.6, pp.197–211, 2023.

6.F. Ahmad, "Scalable Anomaly Detection in Multi-Cloud Environments," Proceedings of the International Conference on Cloud Computing, 2023.

7.J. Yoon, "Zero-Day Attack Detection with Generative AI," IEEE Security and Privacy, vol. 17, no. 2, pp. 97–110, 2023.

8.A. Tabbassum et al., "The Use of Block chain to Enhance Transparency and Accountability in SRE Workflows," 2024 2nd International Conference on Artificial Intelligence, Block chain, and Internet of Things (AIBThings), Mt Pleasant, MI, USA, 2024, pp. 1–6.DOI: 10.1109/AIBThings63359.2024.10763057.

9.A. Tabbassum et al., "Integrating Site Reliability Engineering Principles with DevSecOps for Enhanced Security Posture," 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA), Pune, India, 2024, pp. 1–6. DOI: 10.1109/ICISAA62375.2024.10727769.

10.A.S.Mohammedetal., "Assessin gthe Vulnerability of Machine Learning Models to Cyber Attacks and Developing Mitigation Strategies,"2024ICISAA,Pune,India,pp.1–5.DOI:10.1109/ICISAA62375.2024.10729091.

11.A. Tabbassum et al., "Leveraging Block chain for Secure and Decentralized Identity Management," 2024 IEEE International Conference on Block chain and Distributed Systems Security (ICBDS), Pune, India, pp. 1–6. DOI: 10.1109/ICBDS61729.2024.10737296.