4th IEEE International Conference on AI in Cybersecurity (ICAIC), 5-7 February 2025, University of Houston, Texas, USA

AI Agents for Cloud Reliability: Autonomous Threat Detection and Mitigation Aligned with Site Reliability Engineering Principles

2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC) | 979-8-3315-1888-2/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICAIC63015.2025.10849322

Mourya Chigurupati
Independent Researcher
Austin, TX, USA
0009-0007-7253-959X

Rajesh Kumar Malviya
Enterprise Architect / Independent Researcher
Frisco, TX, USA
009-0003-9831-9190

Arvind Reddy Toorpu
Independent Researcher
Omaha, NE, USA
0009-0003-2560-1523

Karanveer Anand
Technical program manager
San Jose , CA, USA
0009-0007-9204-5490

Abstract— **Cloud environments face constant cybersecurity threats, requiring efficient and reliable defense mechanisms to ensure service continuity. This paper presents a novel approach leveraging AI agents to autonomously detect and mitigate threats in cloud systems, aligned with Site Reliability Engineering (SRE) practices. Using the NSL-KDD dataset, we train AI models to classify attack types accurately, achieving high precision and recall, particularly with Random Forest classifiers. We further simulate threat scenarios to measure the AI agent's response times, comparing Time to Detect (TTD) and Time to Mitigate (TTM) against traditional methods. Results demonstrate a significant reduction in TTD and TTM, with the AI agent achieving up to 6x faster detection and mitigation. This autonomous capability not only improves threat response but also supports SRE-aligned metrics such as Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR), ensuring enhanced reliability in cloud infrastructures. By integrating AI-driven automation into cloud security operations, our findings underscore the potential of AI agents as proactive security operators, advancing both cybersecurity and operational resilience. This research contributes to the development of scalable, autonomous solutions crucial for the future of secure, resilient cloud computing.**

Keywords— AI agents, Cloud security, Threat mitigation, Site Reliability Engineering (SRE), Threat detection

I. INTRODUCTION (*HEADING 1*)

In recent years, cloud infrastructures have become integral to organizations worldwide, providing scalable and flexible solutions for data storage, processing, and application deployment. However, this rapid adoption has also made cloud environments prime targets for increasingly sophisticated cyber threats, including Distributed Denial of Service (DDoS) attacks, brute force intrusions, and advanced persistent threats (APTs) [1]. The dynamic nature of cloud systems, combined with the volume and complexity of modern threats, presents substantial challenges in ensuring continuous security and reliability.

Traditional security approaches often rely on human intervention, which can be time-consuming and less effective in dynamic, high-scale environments. As cloud services expand, manual responses to security incidents are neither efficient nor scalable, underscoring the need for automated solutions. In this context, the emergence of Artificial Intelligence (AI) has opened new possibilities for cybersecurity, enabling faster, more accurate detection and response to threats. While AI has shown promise in various applications, current research primarily focuses on anomaly detection and classification, with limited exploration into autonomous threat mitigation and its integration with Site Reliability Engineering (SRE) principles [2], [3].

AI agents offer a promising solution by autonomously managing security operations within cloud infrastructures. These agents can detect, classify, and respond to threats with minimal human intervention, aligning with SRE practices that emphasize automation, reliability, and efficiency in maintaining service continuity. Autonomous security management through AI agents can significantly reduce Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR) incidents, thus enhancing cloud systems' reliability [4]. Despite these advantages, the practical application of AI agents for real-time detection and mitigation of threats in cloud environments remains underexplored.

This paper proposes a novel framework for leveraging AI agents to enhance cloud security through autonomous threat detection and mitigation, specifically aligned with SRE metrics. Using the NSL-KDD dataset, we develop and evaluate multiple classification models, including Decision Trees and Random Forest classifiers, to categorize attack types. Furthermore, we simulate threat scenarios to assess the AI agent's response capabilities in comparison with traditional human intervention. Results from the simulation demonstrate that AI agents can achieve up to six times faster detection and mitigation, showcasing their potential as proactive security operators within cloud infrastructures. The main contributions of this paper are as follows: (1) Development and evaluation of AI-based classification models for detecting and categorizing cyber threats in cloud systems, with an emphasis on accuracy and efficiency, (2) A simulated comparison of AI agent performance in threat detection and mitigation relative to traditional security approaches, highlighting improvements in SRE-aligned metrics such as MTTD and MTTR, (3) Insights into the integration of AI agents with SRE practices to create a resilient, scalable security framework for cloud environments.

II. RELATED WORK

Recent advancements in artificial intelligence (AI) have opened new pathways for improving cybersecurity,

979-8-3315-1888-2/25/$31.00 ©2025 IEEE

particularly in cloud environments where the volume, velocity, and variety of data make traditional security approaches less effective. Machine learning and deep learning models have shown promise in identifying and responding to various cyber threats, from anomaly detection to real-time attack classification [5]. Supervised learning techniques, for instance, are commonly applied to classify network traffic and detect malicious activities based on labeled datasets like NSL-KDD, UNSW-NB15, and CICIDS2017, which provide structured environments for training and validating AI models in cybersecurity [6]. These datasets allow models to learn patterns in network traffic and distinguish between benign and malicious activities, forming the backbone of AI-driven detection systems.

While AI models excel at anomaly detection, they often fall short in autonomous threat mitigation, a gap that is being addressed through reinforcement learning and AI agents. Autonomous agents, designed to operate with minimal human intervention, have recently gained attention for their potential to automate threat response in cloud infrastructures [7]. For instance, reinforcement learning algorithms are being developed to dynamically adjust security policies based on detected threats, providing a more adaptive defense mechanism [8]. This aligns with Site Reliability Engineering (SRE) practices, where automation and proactive responses are key to maintaining system reliability. Incorporating SRE principles into cybersecurity, however, remains a relatively new research area, with few works focusing on how AI can be optimized to support SRE metrics like Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR) [8].

Another line of research explores hybrid models that integrate both anomaly detection and response mechanisms. AI systems combining supervised classification with unsupervised clustering, for instance, have demonstrated improved detection accuracy by leveraging clustering to identify new attack types, while classifiers handle known threats. The integration of autonomous AI agents for both detection and mitigation represents a significant step forward, as highlighted in recent studies emphasizing the need for proactive security approaches in scalable environments like cloud infrastructures [9].

The growing adoption of cloud-native architectures further amplifies the need for AI-driven security solutions that are both scalable and autonomous. Cloud-native systems are inherently distributed and dynamic, making them susceptible to sophisticated attacks that adapt over time. Recent research indicates that AI can enhance cloud security by applying real-time monitoring and adaptive threat response, ensuring both reliability and resiliency [10]. Yet, while promising, these approaches often lack practical frameworks for deployment in live environments, where the rapid detection and mitigation of threats are essential to prevent significant downtime or data loss.

Building upon these findings, our work addresses the gap in autonomous threat mitigation for cloud security by proposing an AI-driven framework that aligns with SRE practices. Specifically, we demonstrate how AI agents can autonomously detect, classify, and respond to threats in real-time, thereby reducing MTTD and MTTR and enhancing overall system reliability. By focusing on both detection and mitigation, our approach seeks to provide a holistic solution that meets the evolving needs of cloud infrastructures.

III. Methodology

This section outlines the methodology used to develop and evaluate AI agents for autonomous threat detection and mitigation in cloud environments. The proposed approach involves three main phases: data preprocessing, model training for attack type classification, and simulation of threat mitigation capabilities.

A. Data Preprocessing

To create an effective AI-driven solution, the NSL-KDD dataset was used, which is widely accepted for testing intrusion detection systems. The NSL-KDD dataset contains labeled samples of normal and malicious network traffic, representing a variety of attacks that cloud environments typically encounter. We began by preprocessing the dataset to ensure suitability for machine learning. This involved encoding categorical features, normalizing numerical data, and splitting the dataset into training and testing sets to prevent overfitting and allow accurate model evaluation. Data normalization was applied to ensure that feature scales did not bias the model, a common challenge when training neural networks on heterogeneous data [11].

B. Model Training for Attack Type Classification

The second phase involved training AI models to classify network traffic as either benign or malicious. We experimented with multiple machine learning algorithms, including Decision Trees, Random Forests, and Neural Networks, to determine the model that provides the highest accuracy, precision, and recall. Given the complexity of cloud environments, we chose the Random Forest classifier as our primary model due to its robustness against overfitting and its ability to handle large feature sets effectively. Recent studies highlight Random Forest's advantages in dealing with the diverse nature of cloud-based network traffic, especially in distinguishing between similar types of cyber threats [12]. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score, which provided insights into the classification effectiveness across different attack types.

C. Threat Mitigation Simulation

In the final phase, we conducted a simulation to evaluate the AI agent's ability to autonomously detect and mitigate threats in real-time. The simulation involved generating random threat scenarios and measuring the AI agent's **Time to Detect (TTD)** and **Time to Mitigate (TTM)** compared to a baseline (traditional human intervention). These metrics reflect the effectiveness of the AI agent in maintaining cloud reliability and minimizing potential downtime caused by attacks. This phase emphasizes the importance of AI agents not only for threat detection but also for real-time response, a key aspect in high-demand cloud infrastructures [13]. The outcomes of the threat mitigation simulation demonstrated a significant reduction in TTD and TTM with the AI agent, thus supporting its role in proactive cloud security.

By integrating classification and mitigation capabilities, our methodology provides a comprehensive approach to enhancing cloud security through autonomous AI agents. The following section presents empirical results from the classification and mitigation experiments, highlighting the efficacy of the proposed framework.

IV. Empirical Analysis and Results

This section presents the empirical results from the AI-driven classification and threat mitigation simulations. The primary focus is on assessing the AI agent's performance in terms of threat detection accuracy, classification efficiency, and mitigation speed.

A. Attack Type Classification

The classification model was developed to distinguish between normal and malicious network traffic using the NSL-KDD dataset. Multiple algorithms, including Decision Trees, Random Forests, and Neural Networks, were evaluated based on accuracy, precision, recall, and F1-score metrics. Among these, the Random Forest classifier demonstrated superior performance, achieving high accuracy and stability in classifying different attack types. Recent research has corroborated the effectiveness of Random Forests in handling complex and imbalanced datasets, as they are capable of minimizing overfitting through an ensemble approach [14]. In the current study, the Random Forest model reached an accuracy of 99.88%, with F1-scores exceeding 99% for both normal and malicious classes, thus confirming its robustness for real-time security applications in cloud environments.

B. Threat Mitigation Simulation

To evaluate the AI agent's response to real-time threats, a threat mitigation simulation was conducted. The simulation involved generating random threat scenarios and calculating the agent's **Time to Detect (TTD)** and **Time to Mitigate (TTM)** metrics. These were compared to a baseline (traditional human intervention) to determine the improvement in response times. The results indicated a significant reduction in TTD and TTM, with the AI agent responding up to six times faster than the baseline, which aligns with findings in similar studies on automated threat mitigation in dynamic infrastructures [15]. This reduction demonstrates the AI agent's effectiveness in minimizing the impact of cyber threats on cloud service availability and reliability.

C. Integration with Site Reliability Engineering (SRE) Metrics

The empirical results reveal that the AI agent's performance aligns with key Site Reliability Engineering (SRE) metrics, particularly Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR). In reducing TTD and TTM, the AI agent supports SRE's goal of maximizing uptime and reliability. Automated incident response, as facilitated by the AI agent, is essential for maintaining service level objectives (SLOs) in cloud environments, where traditional methods often fall short due to response latency [16]. These findings underscore the role of AI in enhancing cloud infrastructure resilience by integrating proactive security with SRE-aligned automation.

V. Discussion

The findings from this study highlight the potential of AI-driven security agents in transforming cloud security operations by autonomously managing both threat detection and mitigation. The significant reduction in Time to Detect (TTD) and Time to Mitigate (TTM) achieved by the AI agent underscores the efficacy of automated, real-time responses, a critical factor in mitigating the impact of attacks on cloud infrastructure. These improvements in detection and mitigation align closely with Site Reliability Engineering (SRE) goals of minimizing service disruptions and maintaining Service Level Objectives (SLOs), providing an adaptive and resilient security layer for cloud environments [17].

Despite the promising results, certain limitations should be considered. The study relied on simulations and controlled datasets, such as NSL-KDD, which may not fully capture the complexity and variability of threats encountered in live cloud environments. Real-world applications would likely involve a broader array of attack vectors, potentially affecting the AI agent's performance. Additionally, the study did not address adversarial threats specifically aimed at undermining AI-driven systems, which have been shown to compromise the robustness of machine learning models in security applications [18]. Addressing these limitations will be crucial for deploying AI agents at scale in production cloud environments.

Broader implications of this research include the potential for integration of hybrid AI models that combine supervised, unsupervised, and reinforcement learning techniques. Such models could enable AI agents to detect novel attack types, dynamically update response strategies, and improve resilience against adversarial threats. By adopting hybrid AI models, cloud providers can achieve a balance between high detection accuracy and adaptability to emerging threats, creating a more comprehensive cybersecurity approach.

Moreover, the integration of these AI-driven solutions into cloud-native systems would align with the broader trend of automation in SRE practices, allowing for continuous learning and adaptation to evolving threat landscapes. The role of AI in cybersecurity is expected to grow as organizations seek scalable, efficient solutions to protect their increasingly complex infrastructures. The advancements demonstrated in this study provide a foundational approach for future research, encouraging further exploration of adaptive, autonomous security solutions in dynamic cloud settings.

VI. Conclusion and Future Scope

This study introduced an AI-driven framework for autonomous threat detection and mitigation in cloud environments, emphasizing the alignment with Site Reliability Engineering (SRE) practices to enhance system resilience and operational continuity. By leveraging a Random Forest classifier for accurate attack type classification and simulating real-time threat mitigation capabilities, the proposed framework demonstrated substantial reductions in Time to Detect (TTD) and Time to Mitigate (TTM) when compared to traditional human intervention. These findings underscore the role of autonomous AI agents as proactive, adaptive security operators capable of maintaining Service Level Objectives (SLOs) in response to an increasingly complex and dynamic cyber threat landscape.

The primary contributions of this research include the development of a robust, automated solution for managing cloud security threats in real time, as well as a systematic evaluation of AI agents' potential to meet critical SRE-aligned metrics such as Mean Time to Detect (MTTD) and Mean Time to Resolve (MTTR). By integrating machine learning with SRE principles, this framework enhances the reliability and security posture of cloud infrastructures, offering a pathway to scalable, automated threat management.

VII. Future Scope

While the results validate the potential of AI agents in cloud security, future research can address several challenges associated with real-world deployment. Extending the framework to incorporate hybrid AI models—combining supervised and unsupervised learning with reinforcement learning—could enhance detection accuracy, especially for emerging and unknown attack types. Furthermore, as adversarial threats remain a significant concern, incorporating adversarial defense mechanisms into AI-driven security models will be essential to ensure robustness against attempts to manipulate or evade detection systems.

Additional future directions include exploring the integration of AI agents with cloud-native orchestration tools, such as Kubernetes, to automate security responses at the container and microservices level. Such integration could enable rapid, decentralized response strategies that align with the scalability demands of modern cloud applications. Moreover, continuous learning mechanisms could be developed to allow AI agents to adapt to evolving threat patterns, ensuring that the security framework remains effective as new vulnerabilities and attack techniques emerge.

The broader implications of this research extend beyond technical advancements, highlighting the potential of autonomous AI-driven security frameworks to transform industry practices in cloud computing. As organizations increasingly adopt cloud solutions, the integration of adaptive AI agents could streamline compliance with regulatory standards and align with evolving cybersecurity frameworks. Additionally, the advancements demonstrated in this study set the stage for further exploration into creating resilient, fully autonomous cybersecurity frameworks that continuously learn and adapt, ultimately providing a sustainable defense against the growing sophistication of cyber threats in cloud environments. These insights lay a foundation for establishing AI-driven agents as integral components of secure, scalable, and reliable cloud infrastructures, emphasizing the critical role of AI in enabling automated, resilient cloud security solutions.

References

[1] M. S. Ali, S. U. Rehman, N. B. Anuar, and A. S. Malik, "Machine learning-based DDoS attack detection in cloud computing," *IEEE Access*, vol. 8, pp. 59419–59429, 2020.

[2] Y. Wang, X. Yuan, and J. Hu, "Autonomous threat response in cloud environments: A review of AI-based security techniques," *IEEE Trans. Cloud Comput.*, vol. 9, no. 3, pp. 953-962, 2021.

[3] L. Gao, J. Lin, and R. Wang, "Improving security reliability in cloud environments through AI-driven anomaly detection," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 4, pp. 4810–4821, 2021.

[4] D. B. Sookhak, M. T. Ahmed, and K. Zhang, "AI-based approaches for enhancing site reliability engineering in cloud infrastructures," *Future Gener. Comput. Syst.*, vol. 121, pp. 102-116, 2021.

[5] N. Kumar and P. Das, "Reinforcement learning for adaptive cybersecurity policy management in cloud environments," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 1021-1030, 2022.

[6] H. K. Patel and V. R. Upadhyay, "Machine learning frameworks for threat detection in cloud environments," *IEEE Cloud Comput.*, vol. 8, no. 1, pp. 36-45, 2021.

[7] K. Thompson and A. Jones, "AI agents for cybersecurity: Autonomous threat mitigation in cloud infrastructures," *IEEE Trans. Cloud Comput.*, vol. 10, no. 2, pp. 129–137, 2022.

[8] S. Wu, M. Tan, and K. Li, "AI-driven threat intelligence for autonomous cloud security," *IEEE Cloud Comput.*, vol. 10, no. 4, pp. 58-65, 2023.

[9] A. M. Khan and F. B. Gupta, "Hybrid learning models for detecting emerging threats in cloud computing," *IEEE Access*, vol. 11, pp. 2345-2355, 2023.

[10] J. Park, L. Xu, and K. Chong, "Adaptive cloud security: Leveraging AI for real-time threat response," *IEEE Trans. Serv. Comput.*, vol. 15, no. 2, pp. 345-357, 2022.

[11] P. Singh and M. Rani, "Data preprocessing techniques for machine learning in cloud security," *IEEE Trans. Big Data*, vol. 9, no. 1, pp. 145–155, 2022.

[12] T. Q. Lee and R. K. Sharma, "Random forest approach in cybersecurity for cloud environments," *IEEE Trans. Cloud Comput.*, vol. 10, no. 1, pp. 225-234, 2023.

[13] M. Chen and Y. Zhang, "AI-driven real-time response for cloud-native cybersecurity applications," *IEEE Access*, vol. 11, pp. 30711-30723, 2023.

[14] F. Nguyen and G. Anderson, "Evaluating ensemble learning for attack detection in cloud networks," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 4501-4511, 2023.

[15] L. Huang and X. Zhou, "Autonomous cybersecurity response systems in cloud environments," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 2, pp. 183-192, 2023.

[16] Y. Kim and J. Zhao, "Aligning AI-driven cloud security with site reliability engineering metrics," *IEEE Trans. Cloud Comput.*, vol. 11, no. 3, pp. 255-263, 2022.

[17] B. Li and H. Tan, "Improving cloud security resilience with AI-based adaptive measures," *IEEE Trans. Dependable Secure Comput.*, vol. 20, no. 3, pp. 247-259, 2023.

[18] Z. Wang and L. Yang, "Mitigating adversarial threats in AI-powered cloud security systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 1764-1773, 2023.