Towards Privacy-Aware Causal Inference in Cloud Security: Detecting Hidden Threat Paths in Multi-Cloud Ecosystems

[1]Zillay Huma, [2]Hadia Azmat

[1]University of Gujrat, Pakistan

[2]University of Lahore, Pakistan

Corresponding Email: www.zillyhuma123@gmail.com

Abstract

The growing adoption of multi-cloud ecosystems has revolutionized the delivery of computing resources, offering flexibility, scalability, and resilience. However, this distributed architecture introduces significant challenges for security monitoring, particularly in detecting stealthy or hidden threat paths that traverse across heterogeneous platforms. Traditional rule-based intrusion detection and anomaly detection frameworks often fail to uncover complex interdependencies across cloud services, leaving organizations vulnerable to advanced persistent threats (APTs). This research proposes a privacy-aware causal inference framework designed to detect hidden threat paths in multi-cloud environments by modeling causal dependencies within anonymized or obfuscated logs. The approach leverages graph-based causal reasoning combined with privacy-preserving techniques to balance security observability with user data confidentiality. Experimental validation conducted using obfuscated CloudTrail and Azure Activity logs demonstrates that the proposed method achieves high detection accuracy while reducing the risk of privacy leakage. Comparative results against conventional anomaly detection methods reveal superior performance in terms of precision, recall, and explainability. The findings highlight the potential of privacy-aware causal inference to transform cloud security by enabling transparent, robust, and accountable threat detection in complex, distributed infrastructures.

Keywords: Causal inference, multi-cloud security, hidden threat detection, privacy-aware analytics, CloudTrail, causal graphs, adversarial resilience

I. Introduction

The adoption of multi-cloud strategies has become a cornerstone of modern enterprise IT infrastructure. Organizations increasingly rely on a combination of cloud providers such as Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) to diversify risks, optimize performance, and avoid vendor lock-in. While this architectural flexibility enhances business resilience, it introduces significant security challenges due to heterogeneous configurations, varied access policies, and disparate monitoring standards. Adversaries exploit these gaps by constructing hidden threat paths that traverse multiple providers, making detection significantly more difficult than within a single-cloud environment.

Traditional security solutions are often designed with isolated architectures in mind and are not equipped to recognize causal relationships spanning across cloud boundaries. For example, an adversary may exploit a misconfigured identity management system in one provider to escalate privileges and then pivot laterally into another provider's storage services. In such cases, conventional anomaly detection systems may treat each event in isolation and fail to connect the dots, leaving the true threat path hidden. Therefore, an approach capable of uncovering causal interdependencies between events is essential[1].

Causal inference offers a promising foundation for addressing this challenge. By focusing not only on correlation but also on causation, causal inference allows security analysts to reason about the underlying mechanisms that drive malicious activities. In multi-cloud ecosystems, this translates into the ability to identify which events or misconfigurations are causally responsible for triggering subsequent attack steps. Yet, a direct application of causal inference raises significant privacy concerns, as it may require extensive access to sensitive logs and data that organizations are reluctant to share.

The dual requirement of effective threat detection and privacy preservation motivates the development of privacy-aware causal inference. In this paradigm, causal models are constructed using obfuscated, anonymized, or privacy-preserving data representations that retain structural

dependencies while masking sensitive details. This balance ensures that collaborative detection efforts across multiple clouds remain feasible without violating compliance and confidentiality obligations[2].

This research explores the feasibility of privacy-aware causal inference in detecting hidden threat paths within multi-cloud environments. By combining causal graph modeling, privacy-preserving transformations, and experimental validation using obfuscated log data, we present a framework that significantly enhances cross-cloud threat detection capabilities. This work contributes to advancing the state of the art in multi-cloud security while addressing the growing demand for transparent and accountable AI-driven defenses.

II. Literature Review

The study of multi-cloud security has attracted growing attention over the past decade as enterprises shift toward distributed computing models. Early research focused primarily on federated identity management and encryption-based solutions to secure data traversing different cloud environments. While effective for protecting static data, these methods proved less capable of addressing the dynamic and evolving nature of cyber threats, especially those that exploit weak inter-cloud relationships. As attackers began to exploit the complexity of multi-cloud infrastructures, researchers emphasized the need for behavioral and anomaly-based detection mechanisms[3].

Causal inference has emerged as a promising tool for analyzing complex system behaviors. Traditional anomaly detection relies heavily on statistical deviations, but these approaches often fail to distinguish between benign anomalies and genuine threats. By contrast, causal inference techniques can capture dependencies and directional influences between events. In the context of cybersecurity, methods such as Granger causality, Bayesian networks, and structural causal models have been applied to identify attack propagation patterns. These approaches provide deeper insight into why an event occurs rather than simply flagging it as unusual[4].

A parallel research trend has centered on privacy-preserving analytics in distributed environments. Differential privacy, secure multiparty computation, and homomorphic encryption have been widely studied to enable collaborative data analysis without compromising sensitive

information. However, most existing works focus on traditional data domains such as healthcare or finance, with limited exploration in cloud security contexts. Integrating these privacy-preserving techniques with causal inference remains a nascent field, particularly in relation to multi-cloud threat detection[5].

Several recent works have proposed causal-based intrusion detection frameworks. For example, studies utilizing CloudTrail data have demonstrated that causal graph modeling can reveal complex attack chains. Yet, these approaches often assume full access to raw logs, which raises concerns regarding user privacy and compliance with regulatory frameworks such as GDPR. Moreover, the scalability of such systems in multi-cloud ecosystems remains largely untested.

This literature gap highlights the need for novel frameworks that not only model causality across heterogeneous cloud services but also integrate privacy-preserving mechanisms. The present research addresses this by proposing a privacy-aware causal inference model tailored to the detection of hidden threat paths in multi-cloud infrastructures. Unlike previous approaches, our framework explicitly balances the competing demands of security transparency and data confidentiality, thus offering a viable solution for real-world deployments[6].

III. Methodology

The proposed framework leverages causal inference techniques combined with privacy-preserving transformations to detect hidden threat paths in multi-cloud ecosystems. At its core, the system constructs causal graphs from cloud activity logs, such as AWS CloudTrail, Azure Activity Logs, and GCP Audit Logs. Events are treated as nodes in the graph, while potential causal dependencies—derived from temporal order, conditional probability, and domain knowledge—are modeled as directed edges. By applying causal discovery algorithms such as the PC algorithm and Granger causality tests, the system identifies pathways that indicate how one event may have triggered another.

To preserve privacy, raw logs undergo obfuscation before causal modeling. Sensitive attributes such as user identifiers, IP addresses, and resource names are anonymized using pseudonymization and hashing techniques. Additionally, structural causal relationships are retained while masking exact values to ensure that the detection process remains robust. This

transformation allows organizations to contribute activity data for collaborative threat detection without revealing sensitive operational details[7].

A major challenge lies in maintaining the balance between detection accuracy and privacy preservation. To address this, we introduce an adaptive privacy layer that dynamically tunes the level of obfuscation based on data sensitivity and the criticality of threat detection. This is achieved through a feedback mechanism that monitors detection performance metrics while enforcing compliance constraints.

The system further incorporates graph-based reasoning for hidden path detection. Using algorithms such as depth-limited search and causal path enumeration, the framework identifies sequences of events that may represent attack chains across multiple clouds. These paths are ranked based on likelihood scores derived from conditional dependencies, enabling analysts to prioritize investigation. Importantly, the framework provides explainable outputs by explicitly showing causal links, which enhances analyst trust compared to black-box anomaly detection systems[8].

To validate the methodology, a prototype system was implemented and tested using synthetic attack scenarios injected into real obfuscated CloudTrail and Azure Activity logs. These scenarios included privilege escalation, cross-cloud lateral movement, and data exfiltration. Performance was evaluated in terms of precision, recall, false positive rate, and privacy leakage risk. The results are discussed in the following section[9].

IV. Experimental Setup and Results

The experimental validation of the proposed framework was conducted on a testbed emulating a multi-cloud environment. Logs were collected from AWS and Azure deployments configured with vulnerable access policies to simulate real-world attack surfaces. Synthetic attack scenarios were introduced, including credential theft on AWS, lateral privilege escalation in Azure, and cross-cloud data exfiltration to GCP. These scenarios were designed to represent sophisticated adversarial behaviors that would typically evade traditional detection systems[10].

The logs were preprocessed through the privacy-aware obfuscation pipeline before being input into the causal inference module. Baseline models for comparison included a traditional statistical anomaly detector and a machine learning classifier trained on raw event sequences. Performance metrics such as precision, recall, F1-score, and detection latency were measured to evaluate the effectiveness of the proposed approach[11].

Results indicated that the privacy-aware causal inference model outperformed baseline methods across all key metrics. Specifically, the framework achieved a detection precision of 92% and recall of 89%, compared to 78% and 74% respectively for the statistical anomaly detector. The false positive rate was significantly lower, at 6% versus 15%, indicating improved robustness. Importantly, privacy leakage was quantified using entropy-based metrics, showing that anonymization reduced identifiable information by 85% without severely degrading detection performance[12].

The explainability of the causal inference framework emerged as a key advantage. Analysts were able to visualize causal paths linking events across clouds, enabling clear identification of threat propagation sequences. For example, the system successfully revealed that a compromised AWS access key led to privilege escalation in Azure, ultimately facilitating cross-cloud data theft. Such causal insights provided actionable intelligence beyond what conventional black-box models offered[13].

While overall performance was promising, some limitations were observed. In cases of highly aggressive obfuscation, detection accuracy dropped slightly, particularly in scenarios requiring fine-grained user attribution. However, the adaptive privacy layer mitigated this by selectively relaxing anonymization for critical attributes. These results highlight the importance of dynamically balancing privacy and accuracy in real-world deployments[14].

V. Discussion

The findings from the experimental evaluation underscore the potential of privacy-aware causal inference for advancing multi-cloud security. Unlike conventional anomaly detection systems, which often produce a high rate of false positives, the causal graph-based approach demonstrated an ability to capture meaningful interdependencies across disparate cloud platforms. This

enabled the identification of hidden threat paths that would otherwise remain undetected in siloed monitoring systems. By reasoning about causation rather than correlation, the framework provided deeper insights into the mechanisms driving malicious activities[15].

A particularly valuable contribution of this research is the integration of privacy preservation with causal modeling. The use of obfuscation and anonymization ensured that sensitive operational data remained protected, a critical requirement for organizations that are bound by strict compliance regulations. This capability is particularly relevant in collaborative security settings, where multiple organizations may wish to share insights without revealing confidential details. The entropy-based analysis confirmed that privacy leakage could be minimized without significantly undermining detection effectiveness[16].

The experimental results also demonstrated the importance of explainability in cybersecurity. Black-box detection systems, while powerful, often suffer from a lack of interpretability, which hinders analyst trust and slows incident response. By contrast, the causal inference framework provided explicit visualizations of attack chains, enabling analysts to quickly understand and act upon threats. This aligns with the growing emphasis on explainable AI in security applications, where transparency is increasingly viewed as essential for operational effectiveness[17].

Nevertheless, the study highlights several challenges that warrant further research. One limitation lies in the computational complexity of causal graph discovery in large-scale, high-volume log data. While optimizations such as pruning and heuristic search were employed, scalability remains an open question for extremely large cloud infrastructures. Additionally, the reliance on obfuscation may limit the ability to detect highly nuanced insider threats, where subtle variations in user behavior are critical[18].

Looking ahead, potential extensions include the incorporation of federated learning to further enhance privacy by enabling decentralized causal modeling without centralizing sensitive logs. Another promising direction is the application of reinforcement learning to dynamically adjust privacy-obfuscation trade-offs based on evolving threat landscapes. By combining these advances, future systems could provide even stronger guarantees of privacy and security in complex multi-cloud ecosystems[19].

VI. Conclusion

This research demonstrates that privacy-aware causal inference provides a powerful foundation for detecting hidden threat paths in multi-cloud ecosystems. By combining causal graph modeling with obfuscation-based privacy preservation, the proposed framework successfully identified complex cross-cloud attack chains while safeguarding sensitive operational data. Experimental validation confirmed its superiority over conventional anomaly detection methods in terms of precision, recall, and explainability, with minimal privacy leakage. Importantly, the system's causal reasoning capabilities enhanced analyst trust and enabled actionable insights into adversarial behaviors. While challenges remain in scalability and fine-grained insider threat detection, the study establishes a viable pathway toward transparent, privacy-respecting, and robust security analytics for multi-cloud infrastructures.

References:

[1] "Cloud Computing Portability and Interoperability : Cloud Portability and Interoperability." http://www.opengroup.org/cloud/cloud_iop/cloud_port.htm (accessed.

[2] "Masters Thesis - Cloud Computing - Rehan Saleem - download." s available to help (Creswell, 2007, p156-157. http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=1764306&fileOId=1764311 (accessed.

[3] S. Achar, "A Comprehensive Study of Current and Future Trends in Cloud Forensics."

[4] J. Barach, "Federated Learning for Privacy-Preserving Employee Performance Analytics," *IEEE Access,* 2025.

[5] B. Ahmed, A. W. Malik, T. Hafeez, and N. Ahmed, "Services and simulation frameworks for vehicular cloud computing: a contemporary survey," *EURASIP Journal on Wireless Communications and Networking,* vol. 2019, no. 1, pp. 1-21, 2019.

[6] J. Asad and N. Mazher, "Load Balancing Protocol for dynamic resource allocation in cloud computing," 2018.

[7] D. Chen and H. Zhao, "Data security and privacy protection issues in cloud computing," in *Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on*, 2012, vol. 1: IEEE, pp. 647-651.

[8] M. B. A. Dhamande and M. A. Sahu, "Cloud Security Tracking, Log Maintenance and Notification System for Net Banking Cloud Applications," 2014.

[9] J. Barach, "Cybersecurity Project Management Failures," *Indexed in,* vol. 38, 2024.

[10] P. Gaona-García, C. E. Montenegro-Marin, J. D. Prieto, and Y. V. Nieto, "Analysis of Security Mechanisms Based on Clusters IoT Environments," *International Journal of Interactive Multimedia and Artificial Inteligence,* vol. 4, no. Special Issue on Advances and Applications in the Internet of Things and Cloud Computing, 2017.

[11] J. Barach, "Towards Zero Trust Security in SDN: A Multi-Layered Defense Strategy," in *Proceedings of the 26th International Conference on Distributed Computing and Networking*, 2025, pp. 331-339.

[12] K. Hwang, S. Kulkareni, and Y. Hu, "Cloud security with virtualized defense and reputation-based trust mangement," in *Dependable, Autonomic and Secure Computing, 2009. DASC'09. Eighth IEEE International Conference on*, 2009: IEEE, pp. 717-722.

[13] J. Barach, "AI-Driven Causal Inference for Cross-Cloud Threat Detection Using Anonymized CloudTrail Logs," in *2025 Conference on Artificial Intelligence x Multimedia (AIxMM)*, 2025: IEEE, pp. 45-50.

[14] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud computing and grid computing 360-degree compared," in *Grid Computing Environments Workshop, 2008. GCE'08*, 2008: Ieee, pp. 1-10.

[15] J. Barach, "Cross-Domain Adversarial Attacks and Robust Defense Mechanisms for Multimodal Neural Networks," in *International Conference on Advanced Network Technologies and Intelligent Computing*, 2024: Springer, pp. 345-362.

[16] S. Zhang, S. Zhang, X. Chen, and X. Huo, "Cloud computing research and development trend," in *Future Networks, 2010. ICFN'10. Second International Conference on*, 2010: IEEE, pp. 93-97.

[17] J. Barach, "Integrating AI and HR Strategies in IT Engineering Projects: A Blueprint for Agile Success," *Emerging Engineering and Mathematics,* pp. 1-13, 2025.

[18] N. Mazher and I. Ashraf, "A Systematic Mapping Study on Cloud Computing Security," *International Journal of Computer Applications,* vol. 89, no. 16, pp. 6-9, 2014.

[19] J. Barach, "Enhancing intrusion detection with CNN attention using NSL-KDD dataset. In 2024 Artificial Intelligence for Business (AIxB)(pp. 15-20)," ed: IEEE, 2024.