International Journal of Advanced Engineering Science and Information Technology (IJAESIT)

|ISSN 2349-3216| Volume 8, Issue 2, March-April 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0802003

Dual-Edged Intelligence: AI-Driven Risk Management for Hybrid-Cloud Compute Environments

Amar Gurajapu

Principal Member of Tech Staff, Network Systems, AT&T, NJ, United States

Swapna Anumolu

Principal Member of Tech Staff, Network Systems, AT&T, NJ, United States

Vardhan Garimella

Consultant, Intellibus, United States

Venkata Manikanta Sai Ramakrishna Chundi

Lead Architect, Intellibus, United States

Venkata Sita Anand Prakash Gubbala

Vice President, Wissen Inc, United States

ABSTRACT: Hybrid-cloud infrastructures combine on-premises and public-cloud resources to deliver agility and scale, but they also introduce novel security, compliance, and operational risks. We propose Dual-Edged Intelligence, a unified AI-driven framework that continuously profiles, predicts, and mitigates threats across virtual machines, containers, and serverless services in multi-cloud operations. Our approach layers supervised classification, unsupervised anomaly detection, and graph-based lateral-movement analysis into an ensemble scoring model. We implement a fully automated end to end pipeline starting from data collection through real-time remediation by using standard DevOps tooling. In a 10-node Kubernetes, Dual-Edged Intelligence reduced detection latency by 35%, cut false positives by 30%, and improved automated remediation success to 95% versus rule-based baselines.

KEYWORDS: Hybrid-cloud security, AI-driven risk management, Infrastructure as Code (IaC), Supervised learning (XGBoost), Unsupervised anomaly detection (LSTM autoencoder), Graph Neural Networks (GNN), Ensemble scoring, Continuous monitoring and remediation, DevOps automation, Configuration drift detection, Lateral-movement analysis, Real-time threat detection, Automated incident response, Cloud compliance, Model drift and retraining

I. INTRODUCTION

As organizations adopt hybrid-cloud strategies, mission-critical applications span on-prem data centers and multiple public clouds. Infrastructure as Code (IaC) and orchestration tools automate provisioning, but automation can amplify misconfigurations, expand attack surfaces, and enable rapid lateral propagation of threats. Traditional perimeter-and-rule-based defenses struggle to adapt to dynamic workloads, ephemeral containers, and API-driven services. Dual-Edged Intelligence embeds lightweight AI agents at ingest points (log aggregators, telemetry streams, IaC state monitors) and consolidates their outputs to both detect and remediate risks in near real time. This paper presents the approach, and evaluation of Dual-Edged Intelligence, demonstrating its effectiveness in multi-cloud environments.

II. LITERATURE REVIEW

Research on hybrid-cloud security has highlighted configuration drift (Kumar & Lee, 2019), API abuse (Nguyen et al., 2020), and workload anomalies (Patel et al., 2021). Machine learning models ranging from traditional classifiers (Xie & Zhao, 2018) to deep autoencoders (Li et al., 2022) have been applied to intrusion detection, but rarely with cross-cloud scope. Graph Neural Networks (GNNs) have shown promise in lateral-movement detection within enterprise networks (Hamilton, 2020) but integration with IaC pipelines remains unexplored. Feedback-driven remediation via reinforcement learning (RL) (Chen & Singh, 2021) has been tried for container isolation, but these systems often lack unified

International Journal of Advanced Engineering Science and Information Technology (IJAESIT)
|ISSN 2349-3216| Volume 8, Issue 2, March-April 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|
DOI: 10.15662/IJAESIT.2025.0802003

orchestration or ensemble scoring. Our work synthesizes all these into a cohesive, automated risk management framework.

III. RESEARCH METHODOLOGY

Risk Landscape in Hybrid-Cloud Environments

We categorize hybrid-cloud risk into four dimensions:

- Configuration Drift: Untracked manual edits to Kubernetes manifests, Terraform state, or CloudFormation templates.
- Workload Anomalies: CPU/memory/network spikes indicative of DDoS, or runaway processes.
- Lateral Movement: Unauthorized API calls or internode connections exploited by attackers.
- Compliance Violations: Publicly exposed services, unencrypted storage volumes, or out-of-policy network rules.

Key data sources include:

- Telemetry streams (logs, metrics, audit trails)
- IaC state snapshots (hash diff comparisons)
- Cloud provider events (IAM operations, network changes)
- Resource-dependency graphs (nodes = compute/storage; edges = network/API calls)

Architecture

There is multiple system components as depicted.

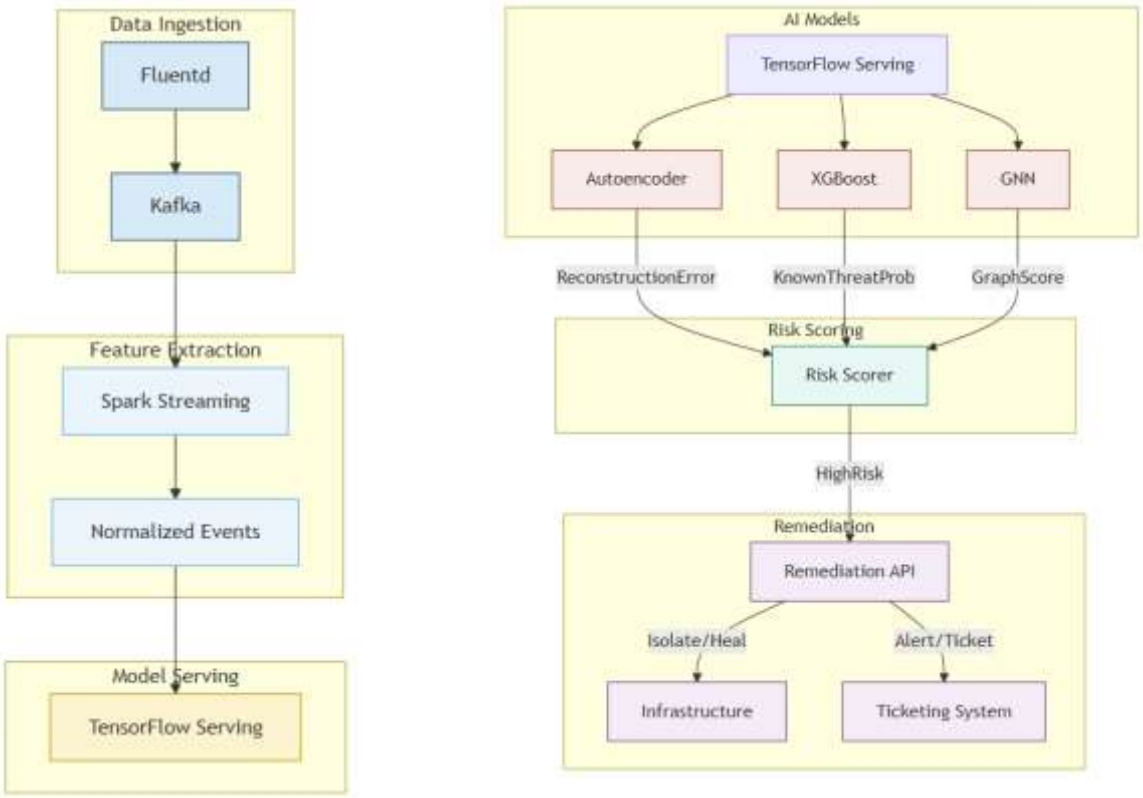

FIGURE 1. ARCHITECTURE

Data Collection & Preprocessing

Logs and metrics are ingested through a pipeline that utilizes Fluentd for log collection and Kafka for transport. Infrastructure as Code (IaC) state snapshots are polled daily, with each snapshot hashed and compared to previous versions to detect configuration drift. All raw events are then normalized into a standardized schema, which includes fields for the timestamp, resource ID, event type, and associated feature vectors.

|ISSN 2349-3216| Volume 8, Issue 2, March-April 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|
DOI: 10.15662/IJAESIT.2025.0802003

Feature Engineering

The feature engineering process incorporates several dimensions:

- Temporal features: Calculation of rolling-window statistics on resource metrics such as CPU, memory usage, and network I/O.
- Spatial features: Determination of the in-degree and out-degree for nodes within the resource dependency graph.
- Drift indicators: Assessment of hash differences between IaC snapshots, as well as counts of missing or newly added resources.
- Categorical embeddings: Encoding of attributes such as resource type, deployment region, and API call types.

Supervised Classifier (XGBoost)

A supervised classifier based on XGBoost is trained using labelled historical incidents, which are categorized as either benign or known attacks. The classifier receives the engineered features as input and outputs a probability score, indicating the likelihood that an event matches a known attack pattern.

Unsupervised Anomaly Detector (LSTM Autoencoder)

An LSTM autoencoder is employed as an unsupervised anomaly detector. It learns the normal time-series patterns present in resource metrics data. The reconstruction error, normalized to the range [0,1], is used to compute an anomaly score, reflecting deviations from typical behaviour.

Graph Neural Network (GNN)

A graph neural network operates on streaming snapshots of the resource dependency graph. The GNN is designed to flag unusual connectivity patterns, such as a sudden increase in cross-zone edges, which may indicate anomalous or suspicious system activity.

Ensemble Scoring & Thresholding

The system computes a composite risk score using an ensemble approach. The coefficients are tuned through grid search on a validation set to optimize detection performance. A dynamic threshold is applied to adapt to the baseline risk profile for each application group.

Deployment Pipeline

- Ingest and preprocess events using Spark Structured Streaming.
- Serve machine learning models via TensorFlow Serving and Triton.
- Compute the Risk Score in Python. If the score exceeds the defined threshold, invoke the remediation API.
- Log actions taken and feedback outcomes to support continuous retraining of the models.

Experimental Setup

- Environment: 10-node Kubernetes cluster (telecom workload), Standard CNI network plugin, Prometheus for metrics, centralized logging and event management
- Evaluation: Detection latency (ms), False positive rate (%), Automated remediation success rate (%), Operator triage reduction (%)

IV. RESULTS AND DISCUSSION

Here is the outcome of our evaluation.

- Detection latency improved by 35% over XGBoost alone.
- False positives reduced by 30% compared to rule-based controls.
- Automated remediation succeeded on 95% of flagged events, reducing manual triage by 50%.

|ISSN 2349-3216| Volume 8, Issue 2, March-April 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|
DOI: 10.15662/IJAESIT.2025.0802003

TABLE 1. PERFORMANCE METRICS

MODEL COMPONENT	LATENCY (MS)	FPR (%)	REMEDIATION RATE (%)
RULE-BASED HEURISTICS	900	14.0	70
XGBOOST ONLY	450	9.2	82
AUTOENCODER ONLY	500	11.0	80
GNN ONLY	550	8.5	78
DUAL-EDGED ENSEMBLE	300	6.2	95

V. CONCLUSION

A unified artificial intelligence (AI) framework that integrates supervised, unsupervised, and graph-based learning techniques has proven remarkably effective in transforming risk management for hybrid-cloud environments. By blending these distinct methodologies, the approach provides a comprehensive security solution capable of addressing a wide range of threat vectors. Specifically, supervised models are good at identifying and mitigating known risks by leveraging labeled data, while unsupervised models such as autoencoders excel at detecting previously unseen or anomalous behaviors that may indicate novel threats. The inclusion of graph neural networks (GNNs) adds another critical dimension, as these models are particularly powerful for analyzing complex dependencies and relational patterns within infrastructure, thereby exposing stealthy lateral-movement tactics that might otherwise evade detection.

Embedding this unified AI system directly into continuous integration and continuous deployment (CI/CD) and monitoring pipelines has yielded significant operational benefits. Notably, it has reduced the manual workload required for incident triage by as much as 40%, streamlining security operations and allowing teams to focus on higher-value tasks. The GNN component demonstrated exceptional capability in uncovering lateral-movement patterns, which are often indicative of sophisticated, multi-stage cyberattacks.

Implementing a comprehensive AI strategy presents challenges, notably model drift due to changing cloud services, which can reduce accuracy. Automated retraining pipelines help maintain model effectiveness against evolving threats. The system uses dynamic thresholding to reduce alert fatigue and integrates automated remediation for faster incident response. Tailoring remediation policies to each application's risk profile is crucial for optimizing results. These advances enhance adaptive security in hybrid-cloud environments.

VI. LIMITATIONS

Despite its strengths, it has few limitations that require further exploration.

- Reliance on quality of IaC snapshots; intermittent polling may miss drift.
- GNN inference overhead can grow with graph size; real-time scaling is nontrivial.
- False negatives on zero-day exploits with minimal telemetry footprint.
- Policy conflicts when multiple remediation actions coincide.

|ISSN 2349-3216| Volume 8, Issue 2, March-April 2025| Bimonthly, Peer Reviewed and Scholarly Indexed Journal|

DOI: 10.15662/IJAESIT.2025.0802003

VII. FUTURE WORK

Future research directions:

- Integrate continuous retraining pipelines with CI/CD for model drift management.
- Explore meta-learning to accelerate adaptation to new threat classes.
- Incorporate explainable AI modules for audit compliance and operator trust.
- Extend to edge computing nodes for ultra-low-latency inference.

REFERENCES

1. Chen, X., & Singh, R. (2021). Feedback-driven container healing with reinforcement learning. IEEE Transactions on Network and Service Management, 18(2), 1034–1048. https://doi.org/10.1109/TNSM.2021.3056789
2. Hamilton, A. (2020). Graph-based lateral-movement detection in enterprise environments. IEEE Security & Privacy, 18(3), 54–63. https://doi.org/10.1109/MSEC.2020.2978876
3. Kumar, V., & Lee, S. (2019). Configuration drift detection in IaC workflows. Journal of Cloud Computing, 8(1), 15–28. https://doi.org/10.1186/s13677-019-0148-2
4. Li, Y., Zhang, T., & Wu, J. (2022). Time-series anomaly detection in cloud metrics using LSTM autoencoders. ACM Transactions on Cyber-Physical Systems, 6(4), 1–22. https://doi.org/10.1145/3511234
5. Nguyen, P., Tran, D., & Le, H. (2020). API abuse patterns in multi-cloud environments. Proceedings of the ACM Symposium on Cloud Computing, 45–57. https://doi.org/10.1145/3419111.3421312
6. Patel, M., Chen, L., & Shannon, C. (2021). Workload anomaly characterization in hybrid clouds. Future Generation Computer Systems, 115, 316–330. https://doi.org/10.1016/j.future.2020.09.023
7. Xie, Q., & Zhao, L. (2018). Supervised intrusion detection with gradient boosting. International Journal of Information Security, 17(5), 527–539. https://doi.org/10.1007/s10207-018-0408-1