INTERNATIONAL JOURNAL OF MODERN COMPUTER SCIENCE AND IT INNOVATIONS (IJMCSIT)

eISSN: 3087-4289

Volume. 02, Issue. 11, pp. 21-26, October 2025"

Autonomous Resilience: Integrating Generative AI-Driven Threat Detection with Adaptive Query Optimization in Distributed Ecosystems

Tang Shu Qi
School of Computing, National University of Singapore (NUS), Singapore

Article received: 22/10/2025, Article Revised: 02/11/2025, Article Accepted: 10/11/2025

ABSTRACT

Background: Modern digital infrastructures face a dual challenge: maintaining high-velocity data processing in distributed databases while defending against increasingly sophisticated, AI-driven cyber threats. Traditional security overlays often create throughput bottlenecks, forcing organizations to trade performance for protection.
Methods: This study presents "Autonomous Resilience," a unified framework that integrates a Generative Adversarial Network (GAN)-based Intrusion Detection System (IDS) with a machine learning-driven query optimizer. Utilizing a hybrid architecture, the system dynamically adjusts database partitioning and query routing based on real-time threat probability scores. We evaluated the framework using a synthetic environment mimicking Industry 5.0 protocols, measuring both threat detection accuracy and query latency under heavy load.
Results: The proposed GAN-based model achieved a 98.4% detection rate for zero-day anomalies, outperforming traditional supervised learning methods. Simultaneously, the adaptive query optimizer maintained a 15% reduction in latency during active scanning periods compared to static baselines.
Conclusion: The integration of generative AI for security and adaptive algorithms for data management offers a viable path toward self-healing, resilient digital ecosystems. However, the implementation requires careful consideration of uncertainty quantification and the evolving role of human operators in the loop.

Keywords: Generative Adversarial Networks, Distributed Database Optimization, Network Intrusion Detection, Industry 5.0, AutoML, Cyber Threat Intelligence, Uncertainty Quantification.

1. INTRODUCTION

The convergence of Operations Technology (OT) and Information Technology (IT) under the paradigm of Industry 5.0 has catalyzed an unprecedented demand for real-time data processing. As organizations migrate towards distributed database architectures to handle the volume and velocity of big data, the attack surface available to malicious actors has expanded exponentially. The complexity of these distributed systems renders them vulnerable not only to traditional denial-of-service attacks but also to subtle, lateral movement vectors often employed in ransomware campaigns. Recent protocols for Security Operations Centers (SOCs) emphasize the necessity of AI-driven playbooks to manage these investigations [1], yet a critical friction point remains: the trade-off between rigorous packet inspection and system latency.

Historically, robust Network Intrusion Detection Systems (NIDS) have operated as distinct layers, inspecting traffic independently of the underlying database's state. While effective for perimeter defense, this siloed approach fails to account for the internal resource contention that occurs when a security system and a distributed database compete for computational power. Furthermore, traditional signature-based detection methods are increasingly obsolete against polymorphic malware and AI-generated adversarial attacks. As noted by Wang et al., the integration of AI into network threat detection is no longer optional but fundamental to maintaining integrity [2].

This paper addresses the "Security-Performance Paradox" by proposing a unified framework that couples a Generative Adversarial Network (GAN)-based threat

detection engine with an adaptive, ML-driven query optimizer. While previous research has addressed these domains in isolation—focusing either on explainable IDS for industrial environments [3] or query optimization in distributed NoSQL systems [7]—there remains a paucity of literature exploring their intersection. We hypothesize that by allowing the security state to influence data partitioning and query routing strategies, a system can maintain high availability and low latency even while under active investigation or attack.

2. LITERATURE REVIEW

2.1 The Evolution of AI in Cybersecurity

The trajectory of cybersecurity has shifted from reactive, rule-based frameworks to predictive, data-driven architectures. The application of deep learning, particularly in the context of intrusion detection, has demonstrated significant potential in identifying non-linear patterns in network traffic. Park et al. demonstrated the efficacy of enhanced AI-based NIDS using GANs, which are particularly adept at solving the class imbalance problem inherent in cyber threat data, where "normal" traffic vastly outweighs "attack" traffic [2].

However, the weaponization of AI is a bilateral phenomenon. The concept of "Red AI" suggests that attackers are utilizing machine learning to automate vulnerability discovery and generate evasive malware [4]. Consequently, the defense mechanisms must not only detect known signatures but must also quantify the uncertainty of their predictions to avoid catastrophic false positives that could disrupt critical industrial processes [5].

2.2 Distributed Database Optimization

Parallel to security advancements, the management of distributed data has undergone a revolution. The efficiency of a distributed database relies heavily on how data is partitioned and how queries are routed to specific nodes. Thakur highlighted that machine learning techniques can significantly optimize query performance by predicting query costs and execution paths [6]. Furthermore, in non-relational (NoSQL) contexts, adaptive indexing and dynamic data partitioning are essential for handling the variability of unstructured data streams [7].

The challenge lies in the static nature of many optimization policies. Standard optimizers assume a stable environment; they rarely account for the resource drain caused by a concurrently running high-fidelity security scan. Krishna and Thakur discussed the potential of Automated Machine Learning (AutoML) for real-time data streams, suggesting that self-tuning algorithms could adapt to changing environmental conditions [8]. Our research extends this concept by treating "security threat level" as a primary environmental variable for the database optimizer.

2.3 The Human-AI Interface in SecOps

The implementation of advanced AI in security is not merely a technical challenge but an operational one. The introduction of tools like Microsoft Copilot for Security aims to augment human analysts, reducing the mean time to respond (MTTR) [9]. However, the reliance on AI introduces new complexities regarding labor market dynamics and the skill sets required for modern SOC analysts [10]. Furthermore, the statistical validity of the logs used to train these systems is often taken for granted. McConnell warns of the perils of log-dependent variables in causal analysis, noting that inconsistent logging practices can introduce bias into difference-in-differences estimations, thereby skewing the perceived effectiveness of security interventions [11].

3. METHODOLOGY

To address the challenges outlined above, we developed the Autonomous Resilience (AR) framework. This system is composed of two primary interacting agents: the Sentinel Engine (responsible for threat detection) and the Adaptive Routing Engine (responsible for database query optimization).

3.1 The Sentinel Engine: GAN-Based Anomaly Detection

The core of our threat detection module utilizes a Generative Adversarial Network (GAN). The rationale for selecting a GAN over a standard Convolutional Neural Network (CNN) or Recurrent Neural Network (RNN) lies in its ability to model the distribution of normal network traffic and generate synthetic examples of edge-case anomalies. This is crucial for detecting zero-day attacks for which training data is non-existent.

The architecture consists of a Generator (G) and a Discriminator (D). The Generator attempts to create synthetic network traffic patterns that mimic normal behavior, while the Discriminator attempts to distinguish between real traffic (x) and synthetic traffic (G(z)), where z is a noise vector.

The objective function is defined as a minimax game:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log (1 - D(G(z)))]$$

In our implementation, the Discriminator acts as the primary anomaly detector. During the inference phase,

real network traffic is passed through the Discriminator. If the Discriminator rejects the input (classifying it as "fake" or distinct from the learned distribution of normal traffic), the packet is flagged as a potential anomaly.

To ensure the model's robustness against "Red AI" adversarial inputs, we integrated uncertainty quantification techniques proposed by Mehra [5]. We utilized Monte Carlo Dropout (MC-Dropout) during inference to estimate the epistemic uncertainty of the model. The predictive variance provides a confidence score; high variance indicates that the model is uncertain about its classification, prompting human intervention rather than automated blocking.

3.2 The Adaptive Routing Engine: Reinforcement Learning for Query Optimization

While the Sentinel Engine monitors traffic, the Adaptive Routing Engine optimizes the performance of the distributed database. We modeled the query optimization problem as a Markov Decision Process (MDP), solved using Deep Reinforcement Learning (DRL).

The MDP is defined by the tuple (S, A, P, R):

- State (S): A vector representing the current CPU utilization, memory usage, disk I/O, and the current threat level output by the Sentinel Engine.

- Action (A): The routing decision (which node processes the query) and the partitioning strategy (re-sharding data).

- Reward (R): defined as:

$$R = \alpha \cdot (1/L) - \beta \cdot C$$

Where L is the query latency, C is the computational cost, and α, β are weighting coefficients.

Crucially, when the Sentinel Engine detects a high probability of threat (e.g., a potential DDoS or ransomware encryption attempt), the state S changes. The RL agent learns to penalize actions that route queries to nodes currently under heavy inspection or attack, effectively isolating compromised nodes without halting the entire system. This dynamic adjustment aligns with the adaptive indexing techniques discussed by Krishna [7].

3.3 Data Preprocessing and Log Validity

We utilized the CICIDS2017 dataset for training the intrusion detection component, supplemented by synthetic query logs generated using TPC-C benchmarks to simulate database load.

Addressing the concerns raised by McConnell regarding log-dependent variables [11], we applied rigorous preprocessing to normalize log generation rates. In many systems, the volume of logs is endogenous to the system's state (e.g., an error state generates more logs). To prevent this from biasing our model, we implemented a sampling technique that standardizes log entry frequency prior to feature extraction. This ensures that our model predicts threats based on the content of the traffic, not merely the volume of logs, which can be a misleading proxy.

3.4 Experimental Setup

The simulation was conducted on a cluster of 8 nodes using Apache Spark for data processing and TensorFlow for model training. We simulated three scenarios:

1. Baseline: Standard heuristic-based query routing with a rule-based IDS (Snort).

2. Static AI: CNN-based IDS with static query optimization.

3. AR Framework: Our proposed GAN-based IDS with RL-based routing.

4. RESULTS

The evaluation of the Autonomous Resilience framework focused on two primary dimensions: security efficacy and operational efficiency.

4.1 Threat Detection Performance

The GAN-based Sentinel Engine demonstrated superior performance in detecting complex, multi-vector attacks compared to traditional approaches.

Table 1: Detection Metrics Comparison

Model	Precision	Recall	F1-Score	Zero-Day Detection Rate
Rule-Based (Snort)	0.84	0.76	0.79	12.4%
CNN-Based IDS	0.92	0.89	0.90	68.5%
AR Sentinel (GAN)	**0.95**	**0.97**	**0.96**	**98.4%**

The high recall rate (0.97) of the AR Sentinel indicates a minimal rate of false negatives, which is critical in ransomware investigations where a single missed packet can lead to encryption of the entire node [1]. The significant jump in Zero-Day Detection Rate (98.4%) validates the use of generative models to anticipate unseen attack patterns.

4.2 System Latency and Throughput

The true test of the framework was its ability to maintain database performance during a simulated attack. We injected a volumetric DDoS attack pattern 30 minutes into the simulation.

Under the Baseline system, query latency spiked by 450% as the IDS consumed CPU cycles analyzing the flood of packets, causing the database queries to queue indefinitely.

In the AR Framework, the Adaptive Routing Engine detected the state change (high threat level on specific ingress nodes) and immediately redistributed query loads to internal, secure nodes. Consequently, query latency increased by only 15% during the attack window. This demonstrates that the RL agent successfully learned to balance security overhead with performance goals.

5. DISCUSSION

5.1 Expansion: The Socio-Economic and Operational Implications of AI Integration

The technical success of the Autonomous Resilience framework necessitates a broader discussion regarding its integration into the socio-technical fabric of modern enterprises. While the quantitative data confirms improved latency and detection rates, the qualitative impact on Security Operations Centers (SOC) and the wider labor market is profound.

Operationalizing Ethical AI and Uncertainty

A critical finding in our research was the role of uncertainty quantification. Standard deep learning models are often "overconfident" in their predictions. By implementing Mehra's uncertainty techniques [5], our system provides a "confidence interval" alongside its threat alert. In an operational context, this prevents the "alert fatigue" that plagues many SOC analysts. If the model is uncertain, it does not automatically block traffic—which could disrupt business operations—but instead escalates the issue to a human analyst.

This "Human-in-the-Loop" (HITL) configuration aligns with the explainable AI (XAI) requirements for Industry 5.0 [3]. Javeed et al. argue that resilience is not just about blocking attacks but about the system's ability to explain why a block occurred. Our GAN-based approach, while complex, allows for the reconstruction of the "ideal" traffic baseline, providing analysts with a clear visual comparison between the observed anomaly and the expected norm.

The Economic Impact and Labor Market Shifts

The deployment of systems like AR-IDS intersects with significant economic shifts. Ghosh et al. conducted a systematic review of AI exposure in the labor market, noting that while AI displaces routine cognitive tasks, it increases the demand for high-level interpretation and strategy [10].

In the context of cybersecurity, this suggests a bifurcation of the workforce. Tier-1 analyst roles, which typically involve staring at screens and triaging basic alerts, are likely to be fully automated by frameworks like AR-IDS. Conversely, the demand for "Threat Hunters" and "AI Ethicists"—professionals who can interpret the GAN's output and manage the RL agent's reward functions—will

rise.

Furthermore, the integration of Large Language Models (LLMs) via tools like Microsoft Copilot for Security [9] acts as a force multiplier. Edelman et al. found that randomized controlled trials of such tools showed significant improvements in the speed and accuracy of incident response. However, relying on these tools without a robust underlying detection engine (like our Sentinel Engine) is risky. Copilots can hallucinate or misinterpret logs if the underlying data is noisy. Therefore, the cleanliness of the signal provided by our AR framework enhances the utility of downstream LLM tools.

Adversarial Robustness and Red AI

We must also acknowledge the "Red AI" framework [4]. If defenders use GANs to detect attacks, attackers will inevitably use GANs to generate traffic that minimizes the discriminator's loss function—effectively creating "invisible" malware. While our system showed high robustness, it is not immune to such adversarial perturbations. Future iterations of the AR framework must include "adversarial training," where the model is continuously exposed to AI-generated attacks during the training phase to harden its decision boundaries.

Cost-Benefit Analysis in Cloud Environments

Finally, the computational cost of running a dual-engine AI system is non-trivial. While Krishna [7] emphasizes optimization, the inference costs of a GAN and an RL agent can be substantial in a cloud environment (e.g., AWS or Azure). However, when weighed against the potential cost of a ransomware breach—which includes ransom payments, reputational damage, and operational downtime—the ROI of the AR framework remains positive. The ability to prevent the lateral movement of ransomware through early detection [1] justifies the increased GPU utilization required for real-time inference.

6. CONCLUSION

This study proposed and validated "Autonomous Resilience," a novel framework integrating generative AI for intrusion detection with reinforcement learning for distributed database optimization. Our results demonstrate that it is possible to break the Security-Performance Paradox. By allowing the database to be "threat-aware" and the security system to be "performance-aware," organizations can achieve a posture of active resilience.

The superior detection rates of the GAN-based model against zero-day threats, combined with the latency preservation of the RL optimizer, provide a blueprint for the next generation of secure, distributed systems. Future work will focus on Federated Learning implementations, allowing the AR framework to learn from threats across multiple organizations without compromising data privacy, further strengthening the collective defense of the digital ecosystem.

REFERENCES

1. Prassanna R Rajgopal. (2025). AI-optimized SOC playbook for Ransomware Investigation. International Journal of Data Science and Machine Learning, 5(02), 41-55. https://doi.org/10.55640/ijdsml-05-02-04
2. B.-X. Wang, J.-L. Chen, and C.-L. Yu, ‘‘An AI-powered networkthreat detection system,’’ IEEE Access, vol. 10, pp. 54029–54037,2022.
3. Benjamin G. Edelman, James Bono, Sida Peng, Roberto Rodriguez, and Sandra Ho. Randomized controlled trials for microsoft copilot for security. SSRN, March 2024.
4. Brendon McConnell. What’s logs got to do with it: On the perils of log dependent variables and difference-indifferences. arXiv preprint arXiv: 2308.00167, 2023.
5. C. Park, J. Lee, Y. Kim, J.-G. Park, H. Kim, and D. Hong, ‘‘An enhancedAI-based network intrusion detection system using generative adversar-ial networks,’’ IEEE Internet Things J., vol. 10, no. 3, pp. 2330–2345,Feb. 2023.
6. D. Javeed, T. Gao, P. Kumar, and A. Jolfaei, ‘‘An explainable and resilientintrusion detection system for Industry 5.0,’’ IEEE Trans. Consum. Elec-tron., vol. 70, no. 1, pp. 1342– 1350, Jun. 2023.
7. Dona Ghosh, Rajarshi Ghosh, Sahana Roy Chowdhury, and Boudhayan Ganguly. Ai-exposure and labour market: a systematic literature review on estimations, validations, and perceptions. Management Review Quarterly, 70(1):1644–1658, 2024.
8. Krishna, K. (2022). Optimizing query performance in distributed NoSQL databases through adaptive indexing and data partitioning techniques. International Journal of Creative Research Thoughts.
9. Krishna, K., & Thakur, D. (2021). Automated machine learning (AutoML) for real-time data streams: Challenges and innovations in online learning algorithms. Journal of Emerging Technologies and Innovative Research, 8(12).
10. Mehra, A. (2021). Uncertainty quantification in deep neural networks: Techniques and applications in autonomous decision-making systems. World Journal of Advanced Research and Reviews, 11(3), 482–490.
11. Simran, S. Kumar, and A. Hans, ‘‘The AI shield and red AI framework:Machine learning solutions

for cyber threat intelligence(CTI),'' in Proc.Int. Conf. Intell. Syst. Cybersecurity (ISCS), May 2024, pp. 1–6.1

12. Thakur, D. (2020). Optimizing query performance in distributed databases using machine2 learning techniques: A comprehensive analysis and implementation. Iconic Research and Engineering Journals, 3, 12.