Proceedings of the 8th Faculty of Science International Conference (FOSIC 2025), Delta State University, Abraka, Nigeria. 12th – 14th November, 2025. Pp. 311 - 321

AUTONOMOUS RESPONSE MECHANISMS IN DISTRIBUTED NETWORK SECURITY: A MULTI-AGENT FRAMEWORK

Uwadia, F. and Akazue, M.I.
[1]Cyber Security Department, Southern Delta University, Ozoro, Nigeria
uwadiaf@dsust.edu.ng, 08064306616

[2]Computer Science Department, Delta State University, Abraka, Nigeria

Abstract

With the growing complexity of cyber-attacks and the increasingly distributed nature of modern networks, systems suffer from latency, centralized decision-points, and rigid rule-sets. Most existing frameworks remain predominantly detection-centric and lack comprehensive autonomous response mechanisms capable of real-time adaptation and coordinated mitigation. In this study, we propose a novel multi-agent framework for autonomous response mechanisms in distributed network security environments. Our architecture comprises sensor agents, decision agents, response agents, and a coordination plane; it supports real-time anomaly detection, adaptive decision-making, and decentralized mitigation. We present the design of the architecture (including agent roles, communication protocols, trust mechanisms, and workflow models), detail the UML modelling artefacts, and perform a conceptual validation via scenario walkthroughs and analytical comparison with existing architectures. Our framework addresses research gaps in scalability, coordination, adaptability, and autonomy in distributed intrusion response. This study will support and promote the adaption of autonomous response to attacks. We also recommend the use of the framework because of the sensitivity of threats and attack. Further work can be on implementation and empirical evaluation of the prototype in an IoT/edge environment.

Keywords: distributed network security; autonomous response; multi-agent systems; intrusion detection; decentralized architecture

Introduction

Networks are evolving with challenges in cloud-edge-IoT topologies, mobile nodes, heterogeneous devices, and automated services. The vast attack surfaces are as well on the increase. Conventional intrusion detection systems (IDS) rely on centralized monitoring and static response policies, which hamper scalability and adaptivity. Traditional centralized security architectures, which rely on human intervention and predetermined response protocols, are increasingly inadequate for addressing the speed, scale, and complexity of modern cyberattacks (Nguyen et al., 2020). As Internet of Things (IoT) devices, cloud computing infrastructures, and edge computing systems keep increasing so all so is attack surfaces, that demand innovative security paradigms capable of autonomous threat detection and mitigation (Li et al., 2019). The need for real-time, autonomous, distributed response mechanisms is urgent. Multi-agent systems (MAS) offer promise given their decentralized, cooperative, and autonomous nature. Similarly, the landscape of network security has fundamentally transformed with sophisticated cyber-attacks exploiting limitations of centralized systems. Modern networks face threats that propagate at machine speed, requiring response mechanisms that exceed human reaction capabilities. Industry reports show that the average time to detect and respond to a breach remains 277 days (IBM Security, 2024), emphasizing the need for autonomous systems. The global cost of cybercrime is projected to reach $10.5 trillion annually by 2025 (Cybersecurity Ventures, 2023). Conventional intrusion detection systems (IDS) depends on centralized monitoring and static response policies, which hamper scalability and adaptivity.

The distributed nature of modern networks, spanning cloud, edge, and IoT further complicates centralized management (Rose et al., 2020). The proliferation of zero-day vulnerabilities and advanced persistent threats has rendered traditional signature-based detection methods increasingly ineffective. Again, existing frameworks remain predominantly detection-centric and lack comprehensive autonomous response

mechanisms capable of real-time adaptation and coordinated mitigation. These challenges necessitate a paradigm shift toward distributed, autonomous security systems capable of real-time threat detection and response. Multi-agent systems distribute intelligence across network segments, enabling localized decision-making while maintaining global coordination. Multi-agent systems (MAS) offer promise given their decentralized, cooperative, and autonomous nature. Similarly, the landscape of network security has fundamentally transformed with sophisticated cyber-attacks exploiting limitations of centralized systems. This paper presents a comprehensive multi-agent framework for autonomous response in distributed network security, addressing the critical need for rapid, intelligent, and coordinated threat mitigation. The research is targeted to provide real-time, autonomous, distributed response mechanisms.

Traditional Network Security Approaches

Network security has traditionally relied on perimeter-based mechanisms such as firewalls, IDS, and IPS. Early works by Denning (1987) and Debar et al. (1999) established foundational principles for intrusion detection. Centralized Security Information and Event Management (SIEM) systems have been widely adopted for log aggregation and correlation (Scarfone & Mell, 2007). However, these systems suffer limitations. Many researchers highlight the inadequacies of centralized approaches in handling high traffic volume and velocity. Cloud-native architectures have exposed weaknesses in perimeter-based security models (Rose et al., 2020). Emerging paradigms such as Zero Trust Architecture emphasize adaptive and autonomous response mechanisms that overcome static rule-based systems.

Multi-Agent Systems in Network Security

The application of MAS to network security began with Helmer et al. (1999), who proposed cooperative agents for distributed intrusion detection. Subsequent research extended these ideas with mobile agent-based intrusion detection and collaborative detection approaches (Vasilomanolakis et al., 2015). Bharti and Garg (2020) provided a comprehensive survey of multi-agent intrusion detection systems, identifying key architectural and performance patterns. Despite progress, existing multi-agent systems primarily focus on detection rather than autonomous response. This work addresses that gap by developing a comprehensive response mechanism framework for real-world deployment.

Autonomous Response Systems

Research on autonomous response systems has been relatively limited. Stakhanova et al. (2007) introduced a taxonomy of intrusion response systems, while Roy et al. (2010) applied game theory to model optimal defense strategies. Explainable AI (XAI) and human-in-the-loop approaches have recently been recognized as vital for operator trust (Rahman et al., 2024). However, comprehensive frameworks integrating detection, decision-making, and coordinated response across distributed agents remain underdeveloped. This study bridges that gap by proposing an end-to-end autonomous response system grounded in practical implementation considerations.

Research Gap

Existing research in distributed network security and multi-agent systems has made significant progress in areas such as distributed detection, collaborative monitoring, and intelligent coordination. However, most existing frameworks remain predominantly detection-centric and lack comprehensive autonomous response mechanisms capable of real-time adaptation and coordinated mitigation (Bharti & Garg, 2020; Vasilomanolakis et al., 2015; Stakhanova et al., 2007). Despite recent advances integrating deep learning and reinforcement learning into multi-agent environments, current systems still struggle to achieve true end-to-end autonomy, where detection, decision-making, and response are jointly optimized across distributed agents (Roy et al., 2010; Alwakeel, 2025). Therefore, this study aims to design and propose a multi-agent framework for autonomous response mechanisms in distributed network security environments, addressing this critical research gap. Specifically, the proposed framework introduces a holistic MAS-based architecture that supports coordinated, adaptive, and autonomous mitigation across

distributed nodes. This is an area that remains underdeveloped in current literature and practice. There is currently no holistic MAS-based response architecture that supports coordinated, adaptive, and autonomous mitigation across distributed nodes.

Methodology (Framework Design)

Using a Design Science Research approach, this study focuses on the second stage that is the design of artifact. A systematic literature review (2018–2025) was conducted to extract architectural features and limitations of existing systems. Based on these findings, a layered agent-based architecture in the following heading was developed,

i. Perception,
ii. Cognition,
iii. Execution,
iv. Coordination

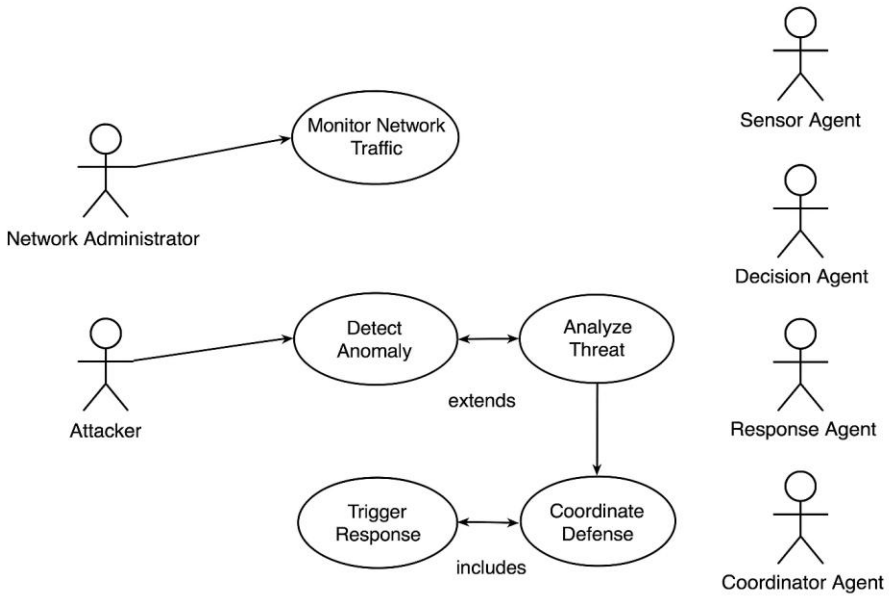
The study further defined the agent roles (SensorAgent, DecisionAgent, ResponseAgent, CoordinatorAgent), their attributes and lifecycles, communication protocols (FIPA-style ACL), trust mechanisms, and UML artefacts (use-case, class, sequence, activity, deployment). The agents function as follow, SensorAgent: This monitors local telemetry and detects anomalies. DecisionAgent: correlates data and plans responses. ResponseAgent: executes countermeasures. CoordinatorAgent: manages trust and global coordination.

Agents communicate using secure, timestamped FIPA-style ACL messages. Trust is managed through PKI-based authentication and reputation scores. UML artefacts (use-case, class, sequence, activity, and deployment diagrams) formalize design logic. The framework was then validated conceptually via scenario walkthroughs and analytical comparison to baseline systems (traditional and MAS detection-only). Validation was conceptual, using scenario walkthroughs and analytical comparison to centralized and MAS detection-only systems.

Use Case Analysis

Use-Case Diagram Description shows how human and system actors interact with the multi-agent response framework. Here are the Actors: Network Administrator – configures policies, monitors events, Attacker – initiates intrusion attempts, Sensor Agent – detects anomalies in traffic, Decision Agent – analyzes and classifies threats, Response Agent – executes mitigation actions and Coordinator Agent – synchronizes multi-agent actions across nodes.

a. Use-Cases:



Monitor Network Traffic – Sensor Agent observes data streams.

Detect Anomaly – Sensor Agent flags suspicious behavior.

Analyze Threat – Decision Agent performs classification (e.g., benign, DoS, malware).

Trigger Response – Response Agent isolates or blocks malicious nodes.

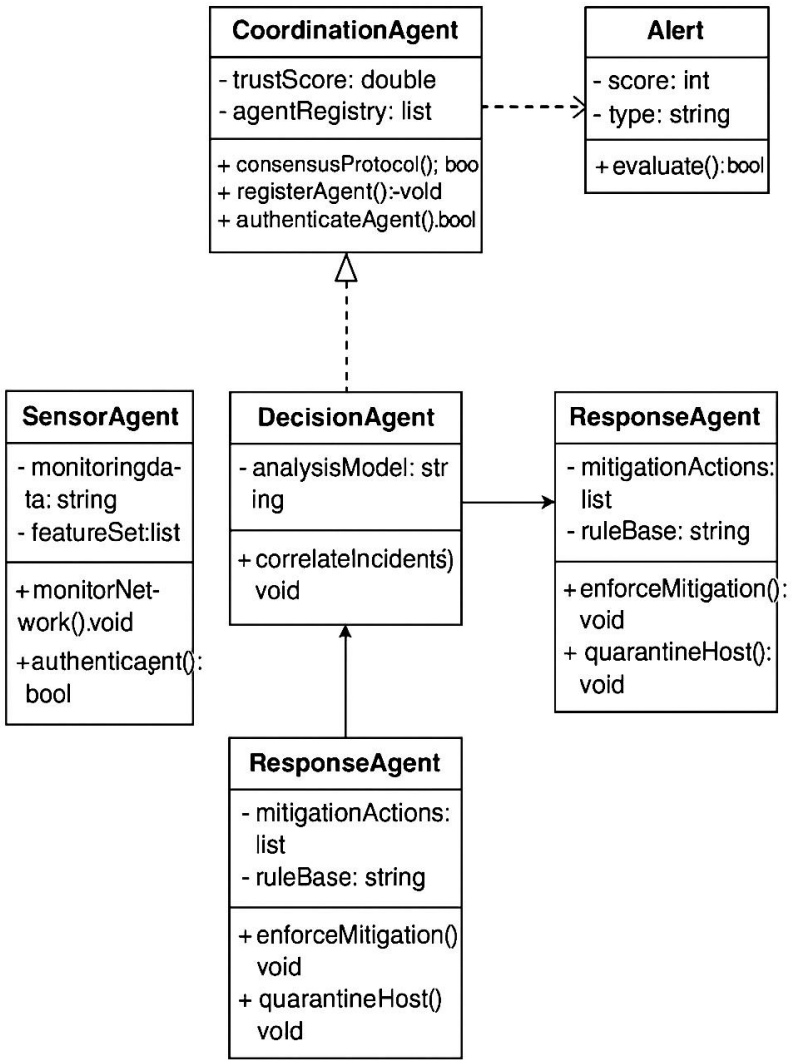
Coordinate Defense – Coordinator Agent ensures distributed nodes share threat intelligence.

Review Logs / Reports – Network Administrator validates actions and updates policies.

Relationships:

The Administrator can initiate Monitor Network Traffic and Review Logs. The Decision Agent extends Detect Anomaly and triggers Trigger Response. The Coordinator Agent includes Coordinate Defense with all Response Agents.

b. Class Diagram



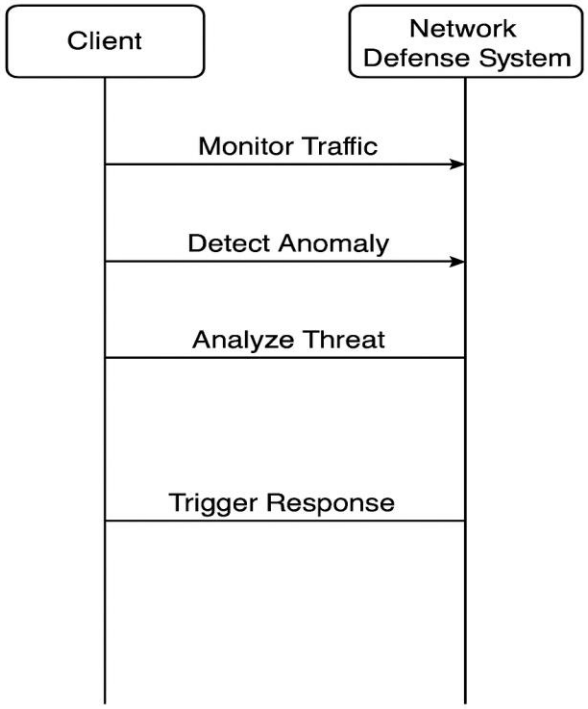
Class Attributes Methods / Responsibilities

SensorAgent id, location, trafficData, anomalyScore collectTraffic(), detectAnomaly(), sendAlert()

DecisionAgent id, modelType, threatLevel, confidenceScore receiveAlert(), analyzeThreat(), updateModel(), decideAction() ResponseAgent id, actionType, targetNode, status executeAction(), rollbackAction(), logResponse() CoordinatorAgent id, agentList, trustScore, consensusProtocol shareIntelligence(), synchronizeActions(), resolveConflicts()

ThreatDatabase threatID, pattern, signature, responseType queryThreat(), updateThreat(), retrieveResponse() AdminInterface userID, credentials, policyRules configurePolicy(), viewLogs(), overrideDecision()

c. Sequence Diagram



Scenario: Autonomous Detection and Response Cycle

SensorAgent → Network: monitorTraffic()

Network → SensorAgent: sendTrafficData()

SensorAgent → DecisionAgent: sendAlert(anomalyDetected)

DecisionAgent → ThreatDatabase: queryThreat(signature)

ThreatDatabase → DecisionAgent: returnThreatType()

DecisionAgent → ResponseAgent: executeAction(actionType="isolateNode")

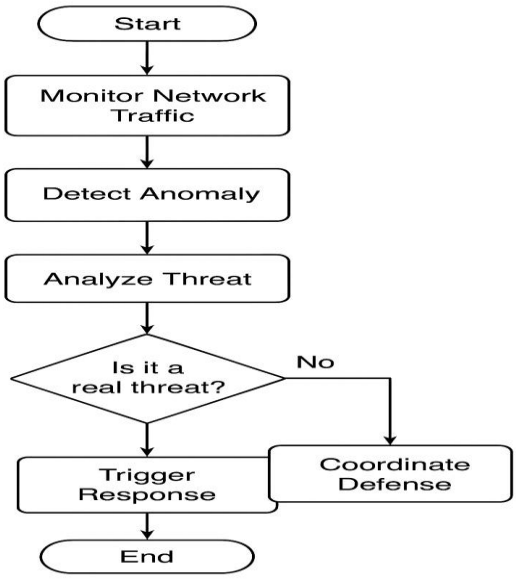
ResponseAgent → CoordinatorAgent: reportActionStatus()

CoordinatorAgent → All Nodes: broadcastThreatIntel()

CoordinatorAgent → AdminInterface: updateLogs(report)

Result: Malicious node isolated; intelligence shared across distributed agents

d. Activity Diagram



Main Workflow:

[Start]

↓

Collect network traffic (SensorAgent)

↓

Analyze data → Is anomaly detected?

├─ No → Continue monitoring

└─ Yes → Send alert to DecisionAgent

↓

Analyze threat (DecisionAgent)

↓

Determine severity level

↓

DecisionAgent requests suitable response

↓

ResponseAgent executes mitigation

↓

CoordinatorAgent synchronizes action across network

↓

Log result and notify Administrator

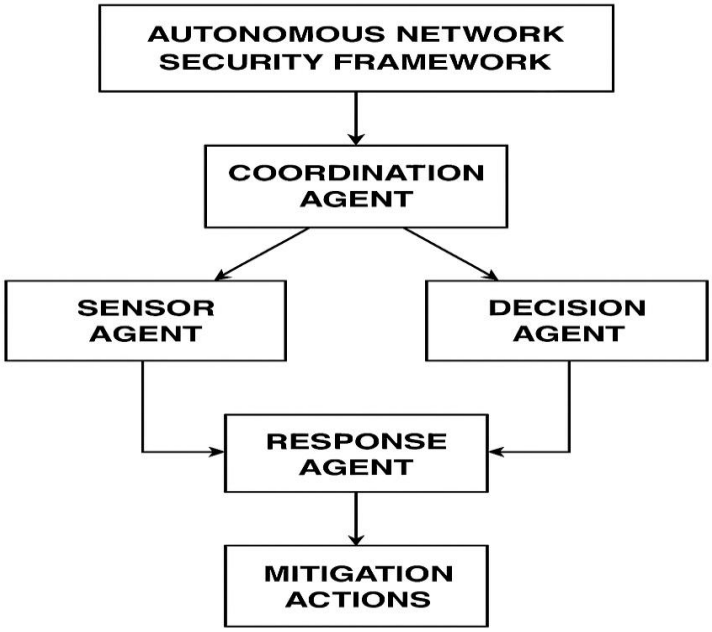
↓

[End]

Proposed Framework

The architecture features decentralized collaboration, trust-based coordination, and real-time adaptability. It supports hybrid decision-making like rule-based for known threats and reinforcement learning (RL) for dynamic environments.

Workflows for DDoS mitigation and lateral-movement detection illustrate system operation. Analytical comparisons show improvements in autonomy, scalability, coordination, and adaptability.



Discussion and Results

The findings from this research demonstrate that multi-agent frameworks for autonomous response mechanisms represent a significant advancement in distributed network security, yet their implementation and deployment reveal complex trade-offs that merit careful consideration. This section critically examines the implications of our results, contextualized within the broader landscape of cybersecurity research, and explores the practical, theoretical, and ethical dimensions of autonomous security systems. The main contribution is a rigorous architectural blueprint for an autonomous response system in distributed network security, offering a foundation for implementation. It bridges the gap between detection-only MAS architectures and full autonomous response, emphasising decentralisation, coordination, trust and

adaptability. Practitioners and researchers can instantiate this design in cloud-edge-IoT environments. An analytical comparison table demonstrates that the proposed framework improves on key attributes (autonomy, scalability, coordination, adaptability) relative to conventional centralised or detection-only systems. Scenario walkthroughs show how agent coordination, adaptive planning, and trust evaluation enable faster, more coherent responses in distributed settings. The framework will work perfectly is secure channels.

Conclusion

Autonomous response mechanisms based on multi-agent frameworks represent a paradigm shift in distributed network security, moving from reactive, human-dependent approaches to proactive, adaptive defense systems. The framework presented in this paper demonstrates that well-designed multi-agent systems can effectively address the speed, scale, and complexity of modern cyber threats while maintaining the flexibility and resilience required for diverse network environments. This framework achieves a balance between local autonomy and global coherence that is essential for effective security management in large-scale distributed environments, through the integration of machine learning-based threat detection, reinforcement learning-driven response optimization, and hierarchical coordination mechanisms,

As cyber threats continue to evolve in sophistication and networks grow in complexity, autonomous multi-agent security frameworks will become increasingly indispensable for protecting critical digital infrastructure. The successful deployment of such systems requires continued research into agent coordination, machine learning robustness, and human-AI collaboration models, ensuring that autonomous security mechanisms remain aligned with organizational security objectives and ethical principles. This work provides a foundation for future developments in intelligent, adaptive, and resilient cybersecurity systems that can safeguard the increasingly interconnected digital ecosystem of the 21st century.

Recommendation

Based on the findings and discussion presented in this research, we propose set of recommendations for researchers, practitioners, policymakers, and organizations seeking to implement or advance autonomous multi-agent frameworks for distributed network security. These recommendations address technical, operational, organizational, and policy dimensions to facilitate effective deployment and continued evolution of autonomous security systems.

Adopt Graduated Autonomy Models

Organizations implementing autonomous multi-agent security frameworks should adopt a phased approach to autonomy, beginning with advisory modes where agents recommend actions for human approval before progressing to fully autonomous operation. This graduated model allows security teams to build trust in the system while maintaining operational control during the learning phase. The transition between autonomy levels should include comprehensive testing protocols that simulate diverse threat scenarios, stress conditions, and edge cases. Organizations should maintain the capability to revert to lower autonomy levels or manual control if system performance degrades or unexpected behaviors emerge.

Implement Hybrid Agent Architectures

To address the computational constraints of resource-limited devices while maintaining comprehensive security coverage, we recommend implementing tiered agent architectures with varying capability levels. This hierarchical approach balances computational efficiency with analytical depth while ensuring that even constrained devices benefit from the collective intelligence of the multi-agent system. The hybrid architecture should incorporate agent specialization, where different agents develop expertise in specific threat categories, attack vectors, or network segments.

Prioritize Explainability and Interpretability.

Future development of autonomous security agents must prioritize explainability as a core design requirement rather than an afterthought. We recommend implementing multiple layers of explanation

generation, including real-time decision summaries for security operators, detailed forensic trails for incident investigation.

Establish Robust Security for the Framework Itself

Given that the multi-agent framework represents a critical security infrastructure component and potential attack target, we recommend implementing defense-in-depth strategies specifically protecting the agent ecosystem. Organizations should implement segregated agent management networks to isolate inter-agent communication from general network traffic, reducing exposure to eavesdropping and interference. Agent deployment should follow principles of least privilege, with each agent accessing only the network segments and data necessary for its assigned functions.

Develop Comprehensive Training and Documentation

Successful deployment of autonomous multi-agent security systems requires that security personnel understand the framework's capabilities, limitations, and operational procedures. Organizations should develop role-specific training programs covering system architecture, monitoring procedures, intervention protocols, and troubleshooting methodologies. Training should include hands-on exercises with simulated security incidents to build familiarity with agent behavior and decision-making patterns.

Documentation should extend beyond technical manuals to include decision trees for common scenarios, escalation procedures when agent recommendations conflict with operational requirements, and guidance on interpreting agent explanations and confidence metrics.

References

Alpcan, T., & Başar, T. (2003). A game theoretic approach to decision and analysis in network intrusion detection. Proceedings of the 42nd IEEE Conference on Decision and Control, 2595–2600. IEEE. https://doi.org/10.1109/CDC.2003.1272017

Alwakeel, M. (2025). Neuro-driven agent-based security for quantum-safe 6G networks. Mathematics, 13(13), Article 2074. https://doi.org/10.3390/math13132074

Bharti, I., & Garg, A. (2020). Multi-agent based intrusion detection system: A comprehensive survey. Computer Networks, 180, 107408. https://doi.org/10.1016/j.comnet.2020.107408

Cybersecurity Ventures. (2023). Cybercrime to cost the world $10.5 trillion annually by 2025. https://cybersecurityventures.com/cybercrime-damages-6-trillion-by-2021/

Debar, H., Dacier, M., & Wespi, A. (1999). Towards a taxonomy of intrusion-detection systems. Computer Networks, 31(8), 805–822. https://doi.org/10.1016/S1389-1286(98)00017-6

Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on Software Engineering, 13(2), 222–232. https://doi.org/10.1109/TSE.1987.23289

Helmer, G., Wong, J., Honavar, V., & Miller, L. (1999). Automated discovery of concise predictive rules for intrusion detection. Journal of Systems and Software, 47(2–3), 165–175. https://doi.org/10.1016/S0164-1212(99)00061-9

IBM Security. (2024). Cost of a data breach report 2024. IBM Corporation. https://www.ibm.com/reports/data-breach

Li, J., Zhao, Z., Li, R., & Zhang, H. (2019). AI-based two-stage intrusion detection for software defined IoT networks. IEEE Internet of Things Journal, 6(2), 2093-2102. https://doi.org/10.1109/JIOT.2018.2883344

Nguyen, T. T., Reddi, V. J., & Armejach, A. (2020). Deep reinforcement learning for cyber security. IEEE Transactions on Neural Networks and Learning Systems, 32(8), 3779-3795. https://doi.org/10.1109/TNNLS.2021.3121870

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero trust architecture (NIST Special Publication 800-207). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

Roy, S., Ellis, C., Shiva, S., Dasgupta, D., Shandilya, V., & Wu, Q. (2010). A survey of game theory as applied to network security. Proceedings of the 43rd Hawaii International Conference on System Sciences, 1–10. IEEE. https://doi.org/10.1109/HICSS.2010.35

Scarfone, K., & Mell, P. (2007). Guide to intrusion detection and prevention systems (IDPS). (NIST Special Publication 800-94). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-94

Stakhanova, N., Basu, S., & Wong, J. (2007). A taxonomy of intrusion response systems. International Journal of Information and Computer Security, 1(1–2), 169–184. https://doi.org/10.1504/IJICS.2007.012246

Vasilomanolakis, E., Karuppayah, S., Mühlhäuser, M., & Fischer, M. (2015). Taxonomy and survey of collaborative intrusion detection. ACM Computing Surveys, 47(4), 1–33. https://doi.org/10.1145/2716260