Intelligent Agents in Multi-Agent Information Security Systems

Roman Yaroviy[1]

[1] *Ivan Franko 1 Private Higher Educational Establishment "European University", 16v. Vernadskoho Akademika Blvd., Kyiv, 03115, Ukraine*

Abstract

In today's digitalization era, the complexity of cyber threats is growing, requiring new approaches to information security. Traditional cybersecurity systems based on monolithic architectures often prove ineffective in IoT environments. Multi-agent systems (MAS), built on the collective interaction of intelligent agents, offer a promising approach to autonomous, adaptive, and scalable protection. This paper examines the use of intelligent agents in multi-agent systems for IoT security, analyzing architectures, key communication protocols, modern use cases, specific challenges, and promising directions for development using artificial intelligence.

Keywords

IoT, intelligent agents, multi-agent systems, artificial intelligence, information security, communication protocols

1. Introduction

The Internet of Things (IoT) encompasses a vast number of resource-constrained devices often connected via unreliable data transmission channels. These devices can serve as points of injection, data redirection, or botnet components. As a result, IoT systems represent high-risk environments where standard centralized security solutions are often cumbersome, slow, or inadequate.

AI-based multi-agent systems — especially those employing reinforcement learning, deep networks, or hybrid methods — can enable IoT systems to:

Detect anomalies and respond locally and autonomously;

Adapt security policies based on network behavior (e.g., device class or risk level);

Scale efficiently as the number of devices grows;

Reduce dependency on centralized servers and lower response latency.

The purpose of this paper is to describe architectures and methods for intelligent agents in IoT security, compare modern approaches, identify challenges and opportunities, and provide recommendations for further research.

2. Concept of Multi-Agent Systems

A multi-agent system is a collection of intelligent agents united by a common goal or task, which they solve through information exchange, coordinated actions, and collective decision-making. Agents in MAS may have different action sets or tasks, and their contexts are not necessarily shared.

[1] Corresponding authors

roman.yaroviy@e-u.edu.ua (R. Yaroviy)

0000-0001-8978-8137 (R. Yaroviy)

2.1. Definition of an Intelligent Agent in IoT

An intelligent agent in an IoT environment is a software component that:

1. Is deployed on a resource-limited device (edge), gateway, or fog node;
2. Can collect data from sensors or network logs;
3. Processes data locally or through shared models;
4. Takes actions (e.g., blocking traffic, isolating devices, changing configurations) and learns from results

2.2. Fundamental Architecture of an Intelligent Agent

The internal structure of an intelligent agent can be viewed as a set of interconnected modules that transform inputs into goal-directed actions.

Core Modules:

- **Decision Integrator**
- **Learning Module**
- **Explainable Engine**
- **Resource Manager**
- **Data Processing Unit**

Operational Flow:
Input Sources → Multilevel Buffering → Data Processing → Core Engine → Multi-Agent System → Output Actions → Environment

Feedback Cycles:
Environment → Result Monitoring → Meta-learning → Model Update → Decision Refinement

Self-Defense System:
Integrity Monitor → Anomaly Detector → Recovery Engine → Blockchain Audit → Security Policy Update

Key Features:

1. **Intelligence:** Adaptive meta-learning, explainability, and quantum-inspired optimization.
2. **Scalability:** Federated learning, neuromorphic computing, and buffered processing under load.
3. **Security & Reliability:** Autonomous defense, blockchain-based auditing, and failure recovery.
4. **User Interaction:** Multimodal interfaces (GUI, API, voice, AR), predictive recommendations, and automated incident documentation.

This architecture enables a transition from reactive to proactive cybersecurity systems with autonomous operation and continuous self-improvement.

3. Future Research Directions

Future development of MAS in cybersecurity focuses on overcoming existing limitations and expanding functionality:

Autonomous SOCs: MAS will underpin self-operating Security Operations Centers (SOCs), where AI handles most operational tasks while humans focus on strategic decisions.

Explainable AI: Developing transparent and trustworthy reasoning mechanisms for regulated sectors.

Enhanced Security Protocols: Protecting agents from manipulation and maintaining the integrity of their decisions.

Standardization: Advancing open protocols (e.g., MCP, A2A) to improve interoperability and reduce vendor lock-in.

Integration with Traditional Systems: Creating hybrid environments by merging MAS with SIEM, SOAR, and TIP platforms.

4. Conclusion

Multi-agent systems based on intelligent agents represent a promising approach to building effective cybersecurity systems capable of countering complex threats. Through autonomy, specialization, and collective coordination, they ensure fast and adaptive responses, proactive threat detection, and efficient security management.
Despite challenges such as coordination complexity, agent vulnerability, and data privacy concerns, the advantages of MAS make them a leading research direction. Further standardization, enhanced security mechanisms, and seamless integration will unlock their full potential for scalable, adaptive, and resilient protection against cyber threats.

References

[1] Alam, Md Morshed; Das, Lokesh Chandra; Roy, Sandip; Shetty, Sachin; Wang, Weichao. "RESTRAIN: Reinforcement Learning-Based Secure Framework for Trigger-Action IoT Environment." *arXiv preprint* arXiv:2503.09513 (2025).

[2] Severt, Marcos; Casado-Vara, Roberto; Martín del Rey, Ángel; Quintián, Héctor; Calvo-Rolle, José Luis. "Multi-agent reinforcement learning based algorithm detection of malware-infected nodes in IoT networks." *Logic Journal of the IGPL*, 2024.

[3] Jamshidia, Saeid; Nikanjama, Amin; Wazed Nafia, Kawser; Khomha, Foutse; Rastab, Rasoul. "Application of Deep Reinforcement Learning for Intrusion Detection in Internet of Things: A Systematic Review." *arXiv preprint* arXiv:2504.14436 (2025).

[4] Andreou, A.; Mavromoustakis, C. X.; Markakis, E.; et al. "Enhancing network slice security with Deep Reinforcement Learning and Moving Target Defense strategies." *Discover Internet of Things* 5 (2025): 67.

[5] Feng, Chao; Huertas Celdran, Alberto; Sanchez Sanchez, Pedro Miguel; Kreischer, Jan; von der Assen, Jan; Bovet, Gerome; Martinez Perez, Gregorio. "CyberForce: A Federated Reinforcement Learning Framework for Malware Mitigation." *arXiv preprint* arXiv:2308.05978 (2023).

[6] Rosenberger, Julia; Urlaub, Michael; Rauterberg, Felix; Lutz, Tina; Selig, Andreas; Bühren, Michael; Schramm, Dieter. "Deep Reinforcement Learning Multi-Agent System for Resource Allocation in Industrial Internet of Things." *Sensors*, vol. 22, no. 11, 2022, article 4099.

[7] Fan, Falong; Li, Xi. "PeerGuard: Defending Multi-Agent Systems Against Backdoor Attacks Through Mutual Reasoning." *arXiv preprint* arXiv:2505.11642 (2025).

[8] "Multi-Agent Deep Reinforcement Learning-based Key Generation for Graph Layer Security." *ACM Transactions on Privacy and Security*, Volume 28, Issue 2, May 2025.

[9] "Multi-Agent Reinforcement Learning for Privacy-Aware Distributed CNN in Heterogeneous IoT Surveillance Systems." *Journal of Network and Computer Applications*, Volume 230, October 2024.

[10] "Multi-Agent Reinforcement Learning for Cybersecurity: Classification and Survey." *Intelligent Systems with Applications*, Volume 26, June 2025