Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

A Review of Security and Privacy in AI-managed Internet of Things Networks

Shiva Ram Linga Rao[1], Vikram Malhothra[2]

[1] Distinguished Systems Engineer, Siemens Cybersecurity. Hyderabad, Telangana, India
[2] Staff Software Engineer, Siemens Cybersecurity. Hyderabad, Telangana, India

Abstract
The Internet of Things (IoT) connects large numbers of devices that are often constrained, long-lived, and weakly secured, which makes IoT networks an attractive target for attackers. As artificial intelligence (AI) is increasingly used to manage these networks, the security picture changes in two ways at once. On one hand, machine learning and, more recently, agentic AI provide stronger tools for detecting and responding to threats. On the other hand, placing AI in control of the network introduces new risks of its own. In this review, the security and privacy of AI-managed IoT networks are examined from both directions. The main IoT threats are summarized, including botnets such as Mirai and the exploitation of known device vulnerabilities. The use of machine learning for intrusion detection is then reviewed, followed by the emerging use of large language model (LLM) based and agentic systems for autonomous defense and management. The new risks that arise when AI is given control, such as model manipulation, unreliable autonomous action, and an enlarged attack surface, are discussed, together with the privacy concerns raised by the large volumes of data these systems collect. Open challenges are identified, and directions for safer AI-managed IoT are suggested. A balanced view of both the promise and the hazards of managing IoT networks with AI is intended to be provided for practitioners and researchers.

Keywords: IoT Security, Intrusion Detection, Agentic AI, Machine Learning, Privacy, Network Management

1. Introduction

The Internet of Things (IoT) has expanded rapidly, connecting billions of devices that range from home sensors and cameras to industrial controllers and medical monitors. Many of these devices are inexpensive and resource-constrained, they are deployed for years without updates, and they are frequently shipped with weak default configurations. As a result, IoT networks present a broad and poorly defended attack surface, and they have become a favoured target for attackers [1], [2].

At the same time, the scale and heterogeneity of IoT deployments have made manual security management impractical. Artificial intelligence is therefore being adopted both to defend these networks and, increasingly, to manage them. This shift has two sides. Machine learning improves the detection of attacks that signature-based tools miss, and newer agentic systems can respond to incidents with little

human involvement. However, once AI is placed in control of a network, the AI itself becomes part of the attack surface, and its failures can have network-wide effects.

In this review, the security and privacy of AI-managed IoT networks are examined from both perspectives. The IoT threat landscape is first summarised. The use of machine learning for intrusion detection is then reviewed, followed by the emerging role of LLM-based and agentic systems. The new risks introduced by AI-based management are discussed next, together with the privacy implications, and the review closes with open challenges and directions for future work.

2. Review Scope and Approach

This work is presented as a narrative review rather than a systematic one. Recent survey articles, peer-reviewed studies, preprints, and industry threat reports were consulted, with an emphasis on material published between 2023 and 2026. The scope is limited to the security and privacy of IoT networks in which AI plays a defensive or managerial role. Topics were selected for their relevance to practitioners, and the material is organised around three questions: what threats IoT networks face, how AI is used to defend and manage them, and what new risks that use introduces.

3. The IoT Threat Landscape

The threats faced by IoT networks are well documented. Weak or default credentials, unpatched firmware, and insecure network services are exploited routinely, and compromised devices are commonly recruited into botnets [2]. The Mirai botnet remains the clearest example. Years after its first appearance, Mirai variants continue to spread by exploiting device vulnerabilities, and recent campaigns have been observed exploiting command-injection flaws in IoT cameras and digital video recorders to build multi-architecture botnets [3], [4].

Beyond botnets, IoT networks are exposed to eavesdropping, spoofing, denial-of-service attacks, and physical tampering. The diversity of protocols and vendors widens the attack surface, because every gateway and translation point is a place where trust boundaries meet. Consumer deployments add a further dimension, since the traffic generated by IoT devices can reveal sensitive information about user behaviour even when it is encrypted [1]. These characteristics make IoT security a moving target that static defences struggle to cover.

4. Machine Learning for IoT Defence

Machine learning has become a central tool for IoT defence, particularly for intrusion detection. Unlike signature-based systems, learning-based detectors can identify previously unseen attacks by modelling normal behaviour and flagging deviations. Classical classifiers such as decision trees, random forests, and k-nearest neighbours have been shown to detect botnet traffic accurately, in some cases directly on resource-constrained devices [5]. Deep learning approaches, including convolutional and recurrent architectures, have been reported to improve detection at the scale of realistic IoT networks [6].

The reported performance of these systems is strong, but comparative studies caution that results depend heavily on the dataset and the evaluation conditions, and that models trained in one environment often transfer poorly to another [7]. Practical intrusion detection for IoT therefore remains constrained by data

quality, the cost of running models on limited hardware, and the need to keep detectors current as attacks evolve.

5. New Risks from AI-managed and Agentic Networks

A more recent development is the use of AI not only to detect attacks but to manage the network itself. Cloud-based frameworks have been proposed in which agents monitor a heterogeneous IoT deployment, configure devices and gateways, and remediate faults with limited human input [8]. LLM-based and agentic approaches extend this further, allowing high-level intents to be translated into autonomous action across the network [9], [10].

This capability introduces risks that did not exist when management was manual. When an AI system is given control, the model becomes a high-value target, and adversarial inputs, data poisoning, and prompt manipulation can cause it to take harmful actions, so that a single compromised agent can affect the whole network [11]. Autonomous action also raises the cost of errors, because a mistaken remediation can propagate faster than a human could intervene. Recent work has begun to address these concerns through design-time safeguards and governance layers that constrain what autonomous systems are permitted to do [12]. The consensus in this emerging literature is that agentic management is promising but must be paired with strong guardrails, human oversight, and auditability.

6. Privacy in AI-managed IoT

Privacy is a distinct concern that AI-based management can both worsen and help. These systems depend on large volumes of telemetry, and centralising that data for training or inference creates a valuable and sensitive store. Even metadata such as traffic timing and volume can expose user behaviour [1]. At the same time, techniques such as on-device inference and federated learning, in which models are trained without moving raw data off the device, offer ways to reduce this exposure. The design choice of where data is processed, at the edge or in the cloud, therefore has direct privacy consequences, and it should be made explicitly rather than by default.

7. Open Challenges and Future Directions

Several challenges remain open. Detection models must be made robust to concept drift and adversarial evasion, because attackers adapt and deployments change over time, and recent work on drift-resilient botnet detection illustrates both the problem and possible responses [13]. The security of the AI management layer itself needs formal treatment, so that guarantees can be made about what an autonomous agent will and will not do. Shared, realistic datasets and benchmarks are needed to make detection results comparable across studies [7]. Finally, privacy-preserving management deserves more attention, so that stronger defence is not achieved at the cost of greater surveillance. Progress is likely to come from combining mature security practice with careful, well-governed use of AI rather than from either alone.

8. Conclusion

The security and privacy of AI-managed IoT networks present a double-edged picture. IoT devices remain weakly protected and heavily targeted, and AI offers real improvements in detecting and responding to threats. However, placing AI in control of the network creates new risks, including model

manipulation, unreliable autonomous action, and privacy exposure from large-scale data collection. The balanced position supported by the current literature is that AI should be adopted for IoT security and management alongside strong guardrails, human oversight, and privacy-preserving design. Treated in this way, AI can strengthen IoT networks; treated carelessly, it can become the weakest link.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

No external funding was received for this work. The authors thank their colleagues for helpful discussions.

References

1. "Analyzing Consumer IoT Traffic from Security and Privacy Perspectives: A Comprehensive Survey", arXiv preprint arXiv:2403.16149, 2024.
2. "Risk and Threat Mitigation Techniques in Internet of Things (IoT) Environments: A Survey", arXiv preprint arXiv:2310.01676, 2023.
3. Akamai Security Intelligence and Response Team, "Here Comes Mirai: IoT Devices RSVP to Active Exploitation", Akamai, 2025, https://www.akamai.com/blog/security-research/active-exploitation-mirai-geovision-iot-botnet.
4. FortiGuard Labs, "Tracking Mirai Variant Nexcorium: A Vulnerability-driven IoT Botnet Campaign", Fortinet, 2025, https://www.fortinet.com/blog/threat-research/tracking-mirai-variant-nexcorium-a-vulnerability-driven-iot-botnet-campaign.
5. "Machine Learning-assisted Intrusion Detection for Enhancing Internet of Things Security", arXiv preprint arXiv:2410.01016, 2024.
6. "Deep Learning-based Intrusion Detection for IoT Networks: A Scalable and Efficient Approach", EURASIP Journal on Information Security, 2025.
7. "Comparative Analysis of Machine Learning based Intrusion Detection in Realistic IoT Networks", arXiv preprint arXiv:2606.31594, 2026.
8. R. Rajidi, "Wi-Fi Evolution and IoT Interoperability Challenges: A Cloud-based Agentic AI Framework for Intelligent Network Management", Proceedings of the 2026 IEEE World AI IoT Congress (AIIoT), 2026, pp. 831-837.
9. "Agentic AI and Large Language Models for Autonomous IoT Cybersecurity: A Systematic Survey, Taxonomy, and Research Roadmap", Electronics, 2026, vol. 15, no. 12, article 2740.
10. "From Intents to Actions: Agentic AI in Autonomous Networks", arXiv preprint arXiv:2602.01271, 2026.
11. "A Survey of Agentic AI and Cybersecurity: Challenges, Opportunities and Use-case Prototypes", arXiv preprint arXiv:2601.05293, 2026.
12. "LLM-driven Approach for Safe and Secure Network Management by Design in IoT-based Systems", Symmetry, 2026, vol. 18, no. 2, article 337.
13. "Concept Drift-resilient IoT Botnet Detection via Latent Space Representation Learning and Alignment", arXiv preprint arXiv:2512.22488, 2025.