

Research Methodology in Computer Science (AI800)

Assignment 1 & 2: Integrated Problem Statement, Research Objectives, and Literature Review

Selected Research Title

Machine Learning-Based Intrusion Detection in Internet of Things Networks

Student ID	262000946
Course Code	AI800
Lecturer	Dr. Mohammed Anbar
Submission Type	Individual Assignment
Date	11 July 2026

1. Introduction

The Internet of Things (IoT) connects smart devices, sensors, gateways, and cloud platforms across smart homes, healthcare, transportation, industrial systems, and other digital environments. This connectivity increases automation and data availability, but it also expands the attack surface because many IoT devices have limited processing power, weak configuration practices, and long deployment lifecycles. These characteristics make IoT networks attractive targets for botnets, denial-of-service attacks, reconnaissance, and data theft.

Traditional signature-based intrusion detection remains useful for known attacks, but it is less effective when attack patterns evolve or when previously unseen attacks appear. For that reason, machine learning (ML) and deep learning (DL) have become important research directions in IoT intrusion detection. Learning-based approaches can identify patterns in normal and malicious traffic and classify suspicious behavior using data-driven models. However, the reviewed literature shows that high detection accuracy alone is not enough for practical IoT security. Effective IoT intrusion detection also requires realistic datasets, lightweight deployment, robustness against changing attack behavior, and protection against adversarial manipulation.

This report builds on the literature review conducted in Assignment 1. It formulates a clear problem statement, defines measurable research objectives, and integrates the reviewed literature to justify the need for further research on machine learning-based intrusion detection in IoT networks.

2. Problem Statement

2.1 Problem Domain

The problem domain of this research is machine learning-based network intrusion detection for IoT environments. IoT networks differ from conventional enterprise networks because devices are often resource-constrained, heterogeneous, continuously connected, and deployed in environments where direct security management may be limited. These characteristics create a need for intrusion detection models that can identify malicious behavior without imposing excessive computational cost on IoT devices or gateways.

2.2 Description of the Problem

Existing studies show that ML and DL models can detect IoT attacks with high reported accuracy. For example, deep learning has been used for distributed attack detection in IoT environments (Diro & Chilamkurti, 2018), autoencoders have been used to detect IoT botnet behavior from compromised devices (Meidan et al., 2018), and supervised classifiers have been evaluated using the BoT-IoT dataset (Ashraf et al., 2025; Koroniotis et al., 2019). However, the main problem is that many proposed IDS models are evaluated mainly on selected datasets or controlled testbeds, and their practical readiness for heterogeneous, resource-constrained, and changing IoT environments remains insufficiently demonstrated.

Several challenges are repeated across the reviewed literature. First, models trained on one dataset may learn dataset-specific patterns rather than general attack behavior, which limits cross-dataset generalization. Second, highly accurate deep learning models may require computational resources that are not always available in edge or gateway-based IoT deployments. Third, adversarial manipulation can reduce the reliability of learning-based IDS models even when their standard accuracy is high (Ibitoye et al., 2019). Fourth, many studies emphasize accuracy, precision, recall, or F1-score, but fewer studies evaluate latency, scalability, explainability, robustness, and deployment cost together.

2.3 Significance of the Problem

This problem is significant because IoT devices increasingly support critical personal, organizational, and industrial activities. If intrusion detection models are accurate only in controlled experiments but unreliable in real deployment, IoT networks may remain vulnerable to botnets, DDoS attacks, reconnaissance, and data theft. In addition, false positives can reduce trust in IDS alerts, while false negatives can allow attacks to continue undetected. Therefore, research is needed to move beyond performance claims on isolated datasets and toward IDS models that are accurate, efficient, robust, and deployable in realistic IoT environments.

2.4 Existing Solutions

The existing solutions in the literature can be grouped into four categories. The first category uses deep learning to learn complex traffic patterns and improve attack classification, as shown by Diro and Chilamkurti (2018) and Ge et al. (2021). The second category uses anomaly detection, such as deep autoencoders, to identify compromised-device behavior without depending only on known attack signatures (Meidan et al., 2018). The third category focuses on realistic dataset development and benchmarking, especially the BoT-IoT dataset, which supports evaluation of IoT-specific attacks (Koroniotis et al., 2019). The fourth category evaluates either robustness or practical classifier choice, such as adversarial robustness testing (Ibitoye et al., 2019) and comparison of supervised classifiers for real-time IoT IDS (Ashraf et al., 2025).

2.5 Justification for Further Research

Although these solutions contribute important knowledge, they do not fully resolve the combined challenge of accuracy, lightweight operation, generalization, and robustness. A practical IoT IDS should not only perform well on one dataset; it should also remain reliable when traffic changes, when attacks are modified, and when deployment resources are limited. Therefore, the research problem can be stated as follows:

Problem Statement: Existing machine learning-based intrusion detection models for IoT networks demonstrate promising detection performance, but there is still a lack of sufficient evidence that these models can simultaneously achieve high accuracy, lightweight deployment, cross-dataset generalization, and adversarial robustness in realistic and resource-constrained IoT environments.

3. Research Goal and Objectives

Primary Research Goal: To propose a lightweight and robust machine learning-based intrusion detection framework for IoT networks that improves attack detection while considering dataset realism, computational efficiency, cross-dataset generalization, and adversarial robustness.

The research objectives required to achieve this goal are:

1. To analyze existing machine learning and deep learning approaches for IoT intrusion detection with specific attention to their datasets, features, evaluation metrics, and deployment assumptions.
2. To identify the main limitations of current IoT IDS studies related to dataset realism, model generalization, computational cost, latency, explainability, and adversarial robustness.
3. To design a conceptual IoT intrusion detection framework that combines effective feature selection, lightweight supervised learning, and deployment at gateway, edge, or fog levels.

4. To define an evaluation strategy for the proposed framework using realistic IoT intrusion datasets and measurable metrics such as accuracy, precision, recall, F1-score, latency, training time, inference time, and robustness under adversarial or changing traffic conditions.
5. To compare the proposed framework conceptually against existing deep learning, anomaly detection, and classical machine learning approaches in terms of detection performance, resource efficiency, and practical deployability.
6. To formulate recommendations for future IoT IDS research that balance model accuracy with operational requirements such as scalability, robustness, and ease of deployment.

4. Literature Review

This section summarizes and critically analyzes six papers relevant to the selected research title. The review focuses on objectives, methods, findings, limitations, and how each paper contributes to the formulation of the research problem.

4.1 Diro and Chilamkurti (2018): Distributed Deep Learning for IoT Attack Detection

Diro and Chilamkurti (2018) proposed a distributed attack detection scheme using deep learning for IoT environments. Their objective was to improve cyberattack detection by using deep learning rather than relying only on shallow learning or centralized detection. The study argues that distributed deep learning is suitable for IoT because attack detection can be performed closer to the network edge, reducing dependence on a single central node.

The strength of this paper is its early focus on distributed detection, which is relevant for edge and fog-based IoT deployments. It also shows the value of automatic feature learning in identifying complex attack behavior. However, the paper is less convincing from a modern deployment perspective because it does not fully address cross-dataset validation, explainability, deployment cost, or adversarial robustness. For this research, the paper is useful because it introduces distributed IDS design as an important direction, but it also highlights the need to evaluate practical constraints more carefully.

4.2 Meidan et al. (2018): N-BalIoT Using Deep Autoencoders

Meidan et al. (2018) introduced N-BalIoT, a network-based anomaly detection method for identifying IoT botnet attacks. The authors extracted behavior snapshots from network traffic and used deep autoencoders to detect anomalous traffic generated by compromised IoT devices. Their evaluation used nine commercial IoT devices infected with Mirai and BASHLITE botnets, which makes the study practically relevant to real IoT attack behavior.

A key strength of this paper is the use of real commercial devices and real botnet families, which improves the realism of the evaluation. The autoencoder approach is suitable for detecting deviations from normal behavior without relying exclusively on predefined signatures. However, the approach is device-specific and depends heavily on stable normal behavior. If device behavior changes over time, the model may produce more false positives. This paper supports the problem statement by showing that realistic device-level evaluation is valuable, but also that generalization remains a challenge.

4.3 Koroniotis et al. (2019): The BoT-IoT Dataset

Koroniotis et al. (2019) developed the BoT-IoT dataset to support research in IoT intrusion detection and network forensics. The authors constructed a testbed that includes normal IoT traffic and multiple attack

categories, including denial-of-service, distributed denial-of-service, reconnaissance, and information theft. The dataset has been widely used because it provides labeled IoT-specific traffic for training and evaluating IDS models.

The contribution of this paper is methodological because dataset quality directly influences IDS evaluation. BoT-IoT allows researchers to compare models on a common benchmark, which is important for reproducibility. However, the dataset is still generated from a controlled testbed, so it cannot represent all real IoT environments. It is also highly imbalanced, which may allow models to report high accuracy while still performing poorly on underrepresented classes. This paper supports the research gap by showing that dataset realism and imbalance must be considered when evaluating IoT IDS performance.

4.4 Ge et al. (2021): Deep Learning-Driven Intrusion Detection for IoT

Ge et al. (2021) proposed a deep learning-driven IDS approach for IoT networks. The authors used packet header field information as generic features and developed feed-forward neural network models with embedding layers to encode high-dimensional categorical variables. The study addressed both binary and multi-class classification tasks and included transfer learning to support feature encoding.

The strength of this paper is its strong methodological design. Packet header fields are practical because they are commonly available in network traffic, and embedding layers help represent categorical traffic features more effectively. The study also moves beyond binary detection by considering multi-class attack classification. However, deep learning models may be computationally expensive for constrained IoT environments. The study would be stronger if it provided deeper evidence about latency, inference cost, and performance under changing attack distributions. This paper supports the need to balance model complexity with deployability.

4.5 Ibitoye, Shafiq, and Matrawy (2019): Adversarial Attacks Against Deep Learning IDS

Ibitoye et al. (2019) examined adversarial attacks against deep learning-based intrusion detection models in IoT networks. The study compared a feed-forward neural network with a self-normalizing neural network using the BoT-IoT dataset. It found that the feed-forward neural network performed better under normal classification metrics, while the self-normalizing neural network showed better resilience against adversarial samples.

This paper is important because it challenges the assumption that high accuracy alone is sufficient. IDS models operate in adversarial settings, where attackers may intentionally manipulate inputs to bypass detection. The paper's limitation is that it focuses on a limited number of models and adversarial techniques. It does not fully evaluate real-time evasion scenarios or defensive strategies such as adversarial training. Nevertheless, it is central to this research because it shows that robustness should be part of IDS evaluation, not an optional additional test.

4.6 Ashraf et al. (2025): Real-Time INIDS with Multiple Machine Learning Classifiers

Ashraf et al. (2025) proposed a real-time IoT network intrusion detection system using the BoT-IoT dataset and multiple supervised machine learning classifiers. The study used preprocessing, feature selection, and classifier comparison involving Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayes, and Artificial Neural Network models. The reported results showed that Random Forest achieved the highest accuracy of 99.2%, while Naive Bayes achieved 98.8%.

The main strength of this paper is its practical comparison of several classifiers using a realistic IoT dataset. It is useful because it shows that classical machine learning can compete strongly with more complex models and may be more suitable for real-time IDS where resource efficiency matters. However, the study still depends mainly on dataset-based evaluation. The high accuracy may be influenced by preprocessing choices, class balancing, or repeated patterns in BoT-IoT. Live deployment testing, cross-dataset validation, and stronger adversarial evaluation would make the results more generalizable. This paper directly supports the research objective of comparing accuracy with operational feasibility.

5. Summary Table of Reviewed Papers

Table 1 summarizes the six reviewed papers according to publication year, research focus, methodology, key findings, and limitations relevant to the problem statement.

Year	Author(s)	Research Focus	Methodology / Dataset	Key Findings	Weaknesses / Limitations
2018	Diro & Chilamkurti	Distributed attack detection for IoT	Deep learning; distributed vs. centralized detection	Distributed deep learning improved attack detection compared with centralized and shallow approaches.	Limited evidence on modern dataset realism, deployment cost, explainability, and generalization.
2018	Meidan et al.	IoT botnet anomaly detection	Deep autoencoders; nine IoT devices infected with Mirai and BASHLITE	Detected compromised-device traffic quickly and accurately.	Device-specific behavior; limited botnet diversity; possible concept drift over time.
2019	Koroniotis et al.	Realistic IoT botnet dataset	BoT-IoT testbed; normal and attack traffic; statistical and ML evaluation	Provided a widely used IoT-specific dataset for IDS and network forensics.	Controlled testbed may not fully represent live IoT environments; severe class imbalance can affect evaluation.
2021	Ge et al.	Deep learning IDS for IoT	Feed-forward neural networks; embeddings; packet header features; binary and multi-class classification	Reported high performance for binary and multi-class IDS tasks.	Computational cost, latency, and changing network conditions require stronger evaluation.
2019	Ibitoye et al.	Adversarial robustness of DL IDS	FNN and SNN tested on BoT-IoT under adversarial samples	FNN performed better in standard metrics; SNN showed better adversarial resilience.	Limited model types and adversarial techniques; no real deployment test.
2025	Ashraf et al.	Real-time IoT NIDS classifier comparison	BoT-IoT; preprocessing; Pearson correlation feature selection; LR, SVM, KNN, DT, RF, NB, ANN	Random Forest achieved the highest accuracy; Naive Bayes was second-best.	Dataset-based evaluation; requires stronger live testing, cross-dataset validation, and robustness checks.

6. Synthesis of the Literature

The reviewed papers are connected by a shared argument: IoT networks require adaptive intrusion detection because static signatures are insufficient against changing and previously unseen attacks. Diro and Chilamkurti (2018), Meidan et al. (2018), Ge et al. (2021), and Ashraf et al. (2025) all support the use of learning-based detection models. Koroniotis et al. (2019) contribute by providing the dataset infrastructure that later models use for evaluation. Ibitoye et al. (2019) extend the discussion by showing that high accuracy must be evaluated alongside adversarial robustness.

A major theme is dataset realism. Meidan et al. (2018) used real commercial IoT devices infected with known botnets, while Koroniotis et al. (2019) developed a structured IoT testbed dataset. Later studies depend on such datasets to evaluate IDS models. This connection shows that dataset quality is not a secondary issue; it directly shapes the credibility of reported model performance.

Another important theme is the tension between deep learning and classical machine learning. Deep learning can learn complex traffic representations, as shown by Diro and Chilamkurti (2018) and Ge et al. (2021). However, Ashraf et al. (2025) show that classical machine learning, especially Random Forest, can achieve very strong detection performance while potentially remaining more practical for real-time deployment. This indicates that the best IDS method is not always the most complex model; it depends on features, computational resources, latency requirements, and deployment context.

The most important research gap concerns practical reliability. Many studies report high accuracy, but fewer studies prove that their models remain effective under cross-dataset testing, adversarial manipulation, or live network deployment. Ibitoye et al. (2019) are especially important because they show that a model can perform well under normal metrics but still be vulnerable to adversarial samples. Therefore, future IoT IDS research should evaluate accuracy together with robustness, explainability, computational efficiency, and deployment feasibility.

7. Conclusion

This report formulated a problem statement and research objectives for the topic Machine Learning-Based Intrusion Detection in Internet of Things Networks. The reviewed literature shows that ML and DL approaches are promising for detecting IoT attacks, but the field still faces important limitations related to dataset realism, model generalization, adversarial robustness, and practical deployment cost. The proposed research direction is therefore justified: IoT intrusion detection research should move toward models that are not only accurate, but also lightweight, explainable, robust, and validated under realistic deployment conditions.

References

- Ashraf, J., Raza, G. M., Kim, B.-S., Wahid, A., & Kim, H.-Y. (2025). Making a real-time IoT network intrusion-detection system (INIDS) using a realistic BoT-IoT dataset with multiple machine-learning classifiers. *Applied Sciences*, 15(4), 2043. <https://doi.org/10.3390/app15042043>
- Diro, A. A., & Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for Internet of Things. *Future Generation Computer Systems*, 82, 761-768. <https://doi.org/10.1016/j.future.2017.08.043>
- Ge, M., Syed, N. F., Fu, X., Baig, Z., & Robles-Kelly, A. (2021). Towards a deep learning-driven intrusion detection approach for Internet of Things. *Computer Networks*, 186, 107784. <https://doi.org/10.1016/j.comnet.2020.107784>
- Ibitoye, O., Shafiq, O., & Matrawy, A. (2019). Analyzing adversarial attacks against deep learning for intrusion detection in IoT networks. arXiv:1905.05137. <https://arxiv.org/abs/1905.05137>
- Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: BoT-IoT dataset. *Future Generation Computer Systems*, 100, 779-796. <https://doi.org/10.1016/j.future.2019.05.041>
- Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Breitenbacher, D., Shabtai, A., & Elovici, Y. (2018). N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, 17(3), 12-22. <https://doi.org/10.1109/MPRV.2018.03367731>