Federated vs. Centralized Malware Detection Robustness on N-BaIoT Under Adversarial Attacks

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the robustness of federated malware detection models against adversarial perturbations compare to centralized models when evaluated on the N-BaIoT dataset with simulated poisoning attacks. This work investigates the possibilities enabled by federated learning concerning IoT malware detection and studies security issues inherent to this new learning paradigm. In this context, a framework that uses federated learning to detect malware affecting IoT devices is. 10 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Federated Learning for Malware Detection in IoT Devices. Research question: How does the robustness of federated malware detection models against adversarial perturbations compare to centralized models when evaluated on the N-BaIoT dataset with simulated poisoning attacks?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

15 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 4.2/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Federated Learning (FL) enables data privacy by design, as data is not shared with any external identity.	×	0.10
FL approaches lack the use of realistic datasets in the FL context, the analysis on adversarial impact, or the discussio	×	0.04
The proposed security framework uses FL to detect, in a privacy-preserving fashion, cyberattacks affecting IoT devices.	✓	0.16
The proposed framework covers both anomaly detection and classification approaches using multi-client FL.	×	0.07
The use case presents a B5G scenario where there is a necessity of detecting cyberattacks affecting IoT devices, managin	×	0.07
The exchange of the model parameters and their aggregation to create a unique and global model can be performed through	×	0.07
After several iterations, each client has a global model obtained as an aggregation of the individual model of each clie	×	0.05
The dataset is split into 79% for training, 1% unused, and 20% for known device test.	×	0.04
The dataset is split into 39.5% for training, 39.5% for threshold selection, 1% unused, and 20% for known device test.	×	0.03
The centralized model achieves 95% accuracy with a 7.8% benign rate and 7% malicious rate.	×	0.04

References

- http://arxiv.org/abs/2104.09994v3
- http://arxiv.org/abs/2512.10637v2
- http://arxiv.org/abs/2006.16545v1