To what extent does the use of differential privacy in federated learning-based malware detection (as seen in

Assignee Research

May 29, 2026

Abstract

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 presented. N-BaIoT, a dataset modeling network traffic of several real IoT devices while affected by malware, has been used to evaluate the proposed framework. Both supervised and unsupervised federated models (multi-layer perceptron and autoencoder) able to detect malware affecting seen and unseen IoT devices of N-BaIoT have

1 Introduction

This paper examines: Federated Learning for Malware Detection in IoT Devices. Research question: To what extent does the use of differential privacy in federated learning-based malware detection (as seen in FEDetect) impact model generalization across different Android API versions, measured by F1-score comparisons on AndroZoo test subsets?.

2 Methodology

Systematic literature search across multiple databases yielded 12 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

12 papers retrieved. 12 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.	×	0.11
Previous works dealing with FL for intrusion detection lack the use of realistic datasets in the FL context, the analysi	×	0.04
The proposed security framework uses FL to detect, in a privacy-preserving fashion, cyberattacks affecting IoT devices.	✓	0.18
The proposed framework covers both anomaly detection and classification approaches using multi-client FL.	×	0.08
The use case presents a B5G scenario where there is a necessity of detecting cyberattacks affecting IoT devices, managin	×	0.08
The model training in FL is performed in a decentralized manner by different nodes, or clients, that use local data.	×	0.07
Each decentralized node trains an individual model using its own data and shares the model parameters (instead of the da	×	0.06
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.	×	0.02
The dataset is split into 39.5% for training, 39.5% for threshold selection, 1% unused, and 20% for known device test.	×	0.01
The centralized model achieves 95% accuracy with a 7.8% benign rate and 7% malicious rate.	×	0.04

References

- http://arxiv.org/abs/2307.01875v1
- http://arxiv.org/abs/2209.14086v2

- http://arxiv.org/abs/2104.09994v3