Unsupervised Federated Learning for Bandwidth-Efficient Malware Detection in Heterogeneous IoT Networks

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the bandwidth utilization of federated learning-based malware detection models scale with increasing device heterogeneity in IoT networks when employing model compression techniques (e.g.,. Federated learning (FL) is an effective paradigm for distributed environments such as the Internet of Things (IoT), where data from diverse devices with varying functionalities remains localized while contributing to a shared global model. By eliminating the need to transmit raw. 12 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: An Efficient Unsupervised Federated Learning Approach for Anomaly Detection in Heterogeneous IoT Networks. Research question: How does the bandwidth utilization of federated learning-based malware detection models scale with increasing device heterogeneity in IoT networks when employing model compression techniques (e.g., quantization, pruning), measured by communication efficiency and detection accuracy trade-offs?.

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 3.5/10.

3 Results

12 papers retrieved. 12 claims extracted; 1 independently verified. Quality review score: 3.5/10.

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
The proposed approach is an unsupervised Federated Learning (FL) method designed for anomaly detection in heterogeneous	✓	0.24
The proposed system consists of four phases: Semantic Data Refinement, Federated Knowledge Aggregation, Intelligent Devi	×	0.07
The Explainable Intelligence Assessment phase applies SHAP to reveal key factors underlying model decisions.	×	0.06
One dataset used in the study is designed for detecting anomalous behaviors, and a second is designed for device identif	×	0.08
Each traffic sample in one of the datasets consists of 46 features serving as input for the autoencoder.	×	0.03
The CICIoT-DIAD 2024 dataset includes a structure file and an HTTP Flood attack file from the DDoS category.	×	0.01
One of the datasets contains 49 features, including 'total length', 'L4 udp', 'L7 http', and 'epoch timestamp'.	×	0.02
One of the datasets contains 47 features, including 'flow duration', 'fin flag number', and 'Target(anomaly detection)'.	×	0.07
One of the datasets contains at least 43 features related to stream jitter, stream counts, and source IP statistics (e.g	×	0.01
Client 1 is associated with the CICIoT2022 Device Identification Dataset.	×	0.09
Client 2 is associated with the CICIoT2023 dataset.	×	0.04
The framework is designed to handle both homogeneous and heterogeneous clients where datasets may differ in feature dime	×	0.03

References

- http://arxiv.org/abs/2104.09994v3
- http://arxiv.org/abs/2602.24209v1

- http://arxiv.org/abs/2507.06449v1