Adaptive vs. Static Weighting in Byzantine-Resilient Personalized Federated Learning for Anomaly Detection

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the comparative effectiveness of adaptive aggregation strategies versus static weighting in mitigating Byzantine failures within personalized federated learning frameworks for network anomaly. Given the distributed nature, detecting and defending against the backdoor attack under federated learning (FL) systems is challenging. In this paper, we observe that the cosine similarity of the last layer's weight between the global model and each local update could be used. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Secure Federated Learning against Model Poisoning Attacks via Client Filtering. Research question: What is the comparative effectiveness of adaptive aggregation strategies versus static weighting in mitigating Byzantine failures within personalized federated learning frameworks for network anomaly detection?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

14 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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
Under the IPM attack with 30% malicious clients and a non-iid degree (q) of 0.1, 0.3, or 0.5 on the MNIST dataset, CosDe	×	0.06
On the MNIST dataset under IPM attack with 30% malicious clients, the No Defense strategy yields a global model accuracy	×	0.08
On the MNIST dataset under IPM attack with 30% malicious clients, the Krum defense strategy yields a global model accura	×	0.08
On the MNIST dataset under IPM attack with 30% malicious clients, the Clipping-Median defense strategy yields a global m	×	0.06
On the MNIST dataset under IPM attack with 30% malicious clients, the CosDefense strategy yields a global model accuracy	×	0.08
On the F-MNIST dataset under IPM attack with 30% malicious clients, the CosDefense strategy yields a global model accura	×	0.08
On the CIFAR-10 dataset under IPM attack with 30% malicious clients, the CosDefense strategy yields a global model accur	×	0.07
In secure aggregation frameworks, individual model updates are locally encrypted and uninspectable by the server.	×	0.07
In secure aggregation frameworks, only the sum of model updates is revealed to the server after sufficient updates are r	×	0.09
Zhao et al. (2020) demonstrated that label information for training data can be computed analytically from the gradients	×	0.07
Empirical results indicate that the last layer's weights are more sensitive to input data distribution compared to other	×	0.04
In an experiment with ten clients training a four-layer CNN independently on a non-iid partitioned MNIST dataset, the av	×	0.11

References

- `http://arxiv.org/abs/2207.09209v4`

- http://arxiv.org/abs/1803.00530v2
- http://arxiv.org/abs/2304.00160v2