Scaling Effects on Inference Efficiency and Certification in Federated Learning Defenses

Assignee Research

May 31, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the effect of model scaling on the inference efficiency and certification bounds of provably secure federated learning defenses against label-flipping attacks. Due to its distributed nature, federated learning is vulnerable to poisoning attacks, in which malicious clients poison the training process via manipulating their local training data and/or local model updates sent to the cloud server, such that the poisoned global model. 8 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FLCert: Provably Secure Federated Learning against Poisoning Attacks. Research question: What is the effect of model scaling on the inference efficiency and certification bounds of provably secure federated learning defenses against label-flipping attacks?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

10 papers retrieved. 8 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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
FLCert is empirically evaluated on five datasets from different domains, including three image classification datasets (	×	0.05
FLCert-D with FedAvg and N = 500 can achieve a certified accuracy of 81% on MNIST when evenly distributing the training	×	0.05
FLCert provides provable security guarantees against poisoning attacks, with FLCert-P offering probabilistic security gu	×	0.15
FLCert is evaluated on multiple datasets from different domains and multiple base FL algorithms, showing that it is secu	×	0.11
Federated Learning (FL) involves n clients and a server, where each client holds a local training dataset and collaborat	×	0.11
FL is vulnerable to poisoning attacks as shown in recent works.	×	0.11
FL has been deployed by industry, such as Google for next-word prediction on Android Gboard.	×	0.02
Existing FL methods mainly follow a single-global-model paradigm, involving iterative communication between clients and	×	0.09

References

- http://arxiv.org/abs/2509.22873v2
- http://arxiv.org/abs/2210.00584v2
- http://arxiv.org/abs/2412.18507v1