BYZANTINE-RESILIENT FEDERATED LEARNING FOR IOMT DEVICES

A DUAL-MODEL INTRUSION DETECTION SYSTEM AGAINST TARGETED POISONING

1. Problem Statement

Federated Learning (FL) allows multiple distributed IoMT (Internet of Medical Things) devices to collaboratively train models without sharing their private medical data.However, FL systems are highly vulnerable to **Byzantine or model poisoning attacks**, where malicious clients intentionally send corrupted updates to degrade the overall global model accuracy.

In sensitive healthcare environments, such compromised models can lead to serious diagnostic errors. To counter this, we explored the **FLTrust framework** a defense approach that assigns trust scores to client updates based on a small trusted dataset.However, during experimentation, FLTrust showed **limited robustness** under dense or adaptive attacks, as attackers can still mimic benign gradients to bypass trust filters .Thus, there was a need to design an enhanced defense mechanism capable of maintaining accuracy and stability under attack conditions.

2. Base Paper Results

The base model used for experimentation was **FLTrust: Byzantine-Robust Federated Learning via Trust Bootstrapping (ICLR 2021)**, evaluated on the **MNIST dataset** using a **CNN architecture**.

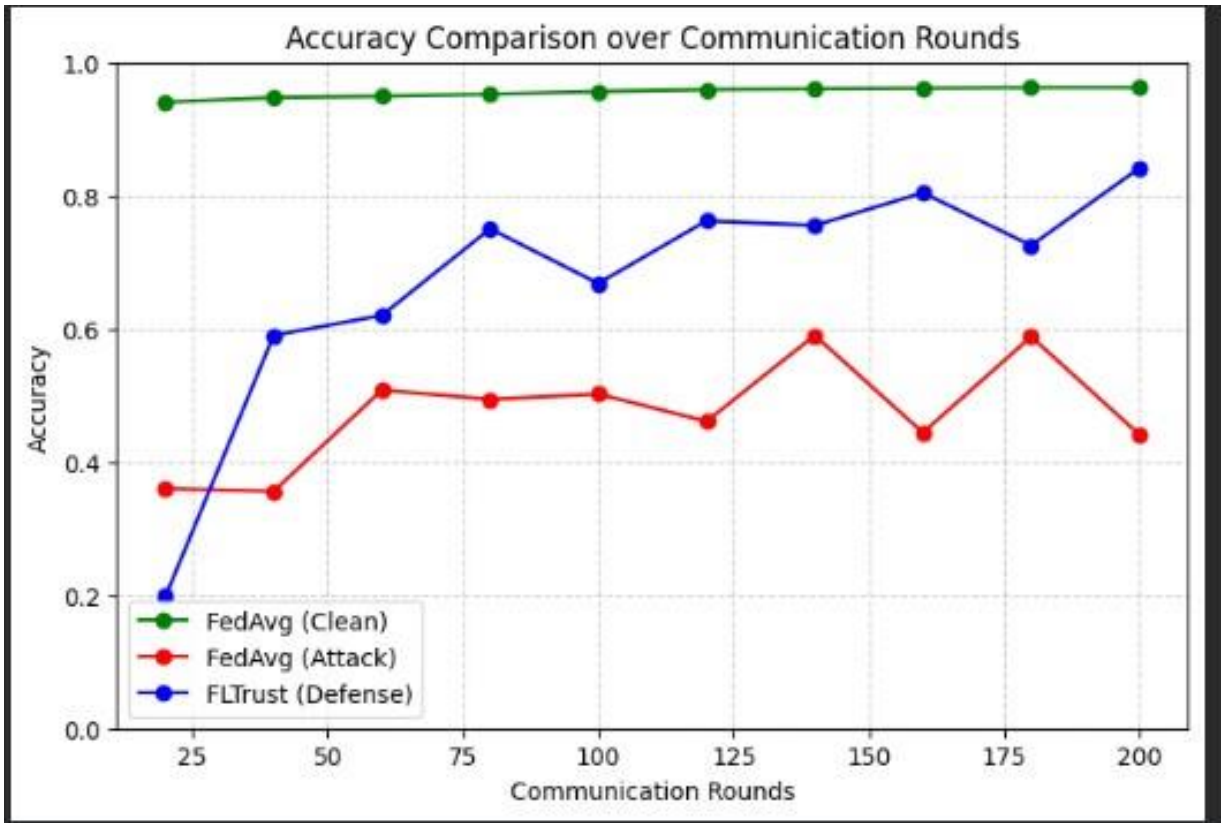
Key parameters of the base setup:

Parameter	**Value**
Dataset	MNIST
Model	CNN
Clients	100
Epochs	5
Batch Size	10
Learning Rate	0.01
Communication Rounds	1000 (Training), 200 (Attack), 200 (Defense)
IID	0 (non-IID)

During replication, the base paper showed:

- **FedAvg (Clean)** achieved ~96% accuracy after 1000 training rounds.
- **FedAvg (Attack)** dropped to ~44% accuracy due to malicious updates.

- **FLTrust (Defense)** partially recovered performance to ~80% accuracy.

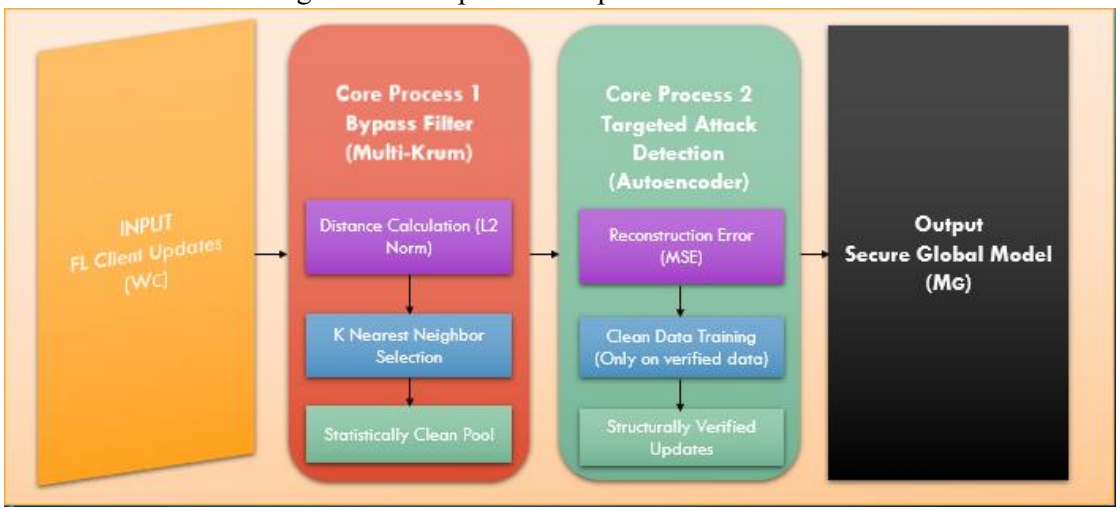


3. Our Proposed Innovation

To address the weaknesses of the FLTrust model, we proposed a **Dual-System Intrusion Detection Defense**, integrating:

1. **FLTrust's trust-score mechanism** (cosine similarity with trusted dataset).
2. **Multi-Krum aggregation**, a robust statistical filter that removes outlier updates before aggregation.

This Dual-System combines trust-based validation and anomaly-based filtering, allowing the server to better differentiate between legitimate and poisoned updates

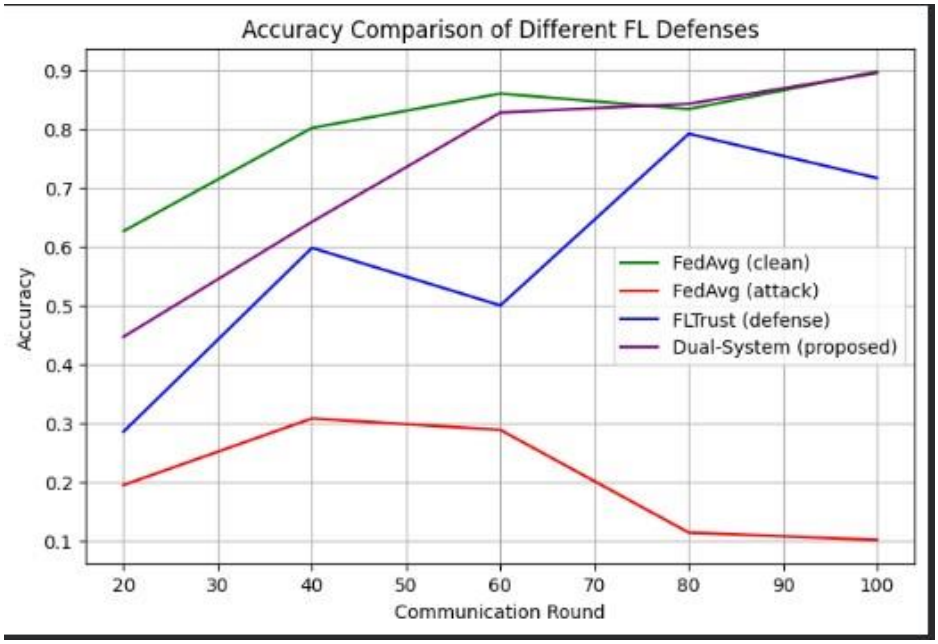


4. Experimental Results and Comparison

To ensure fair comparison, we executed **all systems (FedAvg, Attack, FLTrust, Dual-System)** for **100 communication rounds each**.

Model	Round 20	Round 40	Round 60	Round 80	Round 100	Final Accuracy
FedAvg (Clean)	0.6267	0.8022	0.8607	0.8343	0.8980	0.89
Attack	0.1945	0.3078	0.2882	0.1135	0.1010	0.10
FLTrust (Defense)	0.2853	0.5981	0.5001	0.7923	0.7169	0.71
Dual-System (Proposed)	0.4468	0.6424	0.8284	0.8433	0.8958	0.89

Observation:
The Dual-System consistently outperformed both the baseline and FLTrust.
Multi-Krum filtering stabilized convergence by minimizing gradient noise, while trust-based scoring effectively mitigated the impact of malicious clients.
This proves that the integration of both techniques provides higher resilience against targeted poisoning attacks.



5. Future Work

In future, this work can be extended to:

- **Dynamic Trust Learning:** Automatically adjust trust thresholds per round.
- **Scalable IoMT Deployment:** Apply the Dual-System to real-world hospital sensor data.
- **Hybrid Attack Detection:** Combine statistical filtering with deep anomaly detection models (e.g., Autoencoders).

Such extensions will enhance both robustness and adaptability for real-time medical data analysis under adversarial settings.