Federated Learning Poisoning Attack (FLPA) represents a critical adversarial machine-learning threat, particularly relevant in **5G and beyond** mobile ecosystems where distributed intelligence is fundamental. Federated Learning (FL) allows multiple edge devices—such as IoT nodes, mobile devices, or base-station-side compute modules—to train models locally and send model updates to a central aggregator. This paradigm aligns with the **edge-computing and AI-native design of 5G/6G networks**, enabling low-latency inference and scalable deployment.

However, FL implicitly trusts client-side computations, creating a broad attack surface.
A poisoning attack occurs when a malicious client manipulates local data or computed model parameters, influencing the global model during aggregation. In next-generation networks, where thousands to millions of edge devices contribute to AI-driven decisions (e.g., traffic prediction, anomaly detection, mobility management, beamforming optimization), even a **single compromised node** can degrade global model integrity or insert targeted backdoors.

In the project's demonstration, benign clients compute Random Forest results on partitioned datasets, while the adversarial client sends corrupted updates or falsified accuracy metrics. The central federated learner computes a metric such as mean accuracy, unknowingly incorporating the poisoned contribution. This reflects realistic challenges faced in 5G/6G environments:

- **Edge devices are often untrusted** and physically exposed
- **Device diversity leads to inconsistent data distributions**, increasing poisoning impact
- **Network slicing isolates traffic, but not model aggregation**
- **Distributed AI in RAN and core functions relies on model integrity**

FL poisoning attacks generally fall into:

Data Poisoning

Injecting mislabeled or crafted samples into local datasets to shift the global model.

Model Poisoning

Manipulating gradients, weights, or reported metrics before uploading them.

Consequences are severe in telecom scenarios:

- **Degraded anomaly detection accuracy** in slice security systems
- **Misclassification of malicious traffic as benign**
- **Backdoors embedded into distributed AI workflows**
- **Compromised orchestration or traffic-engineering decisions**
- **Propagation of poisoned model parameters across edge clouds**

As 6G research moves toward AI-native networking, FLPA underscores the need for:

- Robust client validation
- Secure aggregation (e.g., Krum, Multi-Krum, Trimmed Mean)
- Behavioral scoring of edge participants
- Cryptographic consistency checks
- Continuous anomaly detection across training rounds

The FLPA case demonstrates that **AI-driven telecom architectures are only as secure as their weakest learning participant**, reinforcing the need for ML-specific security controls across the entire 5G/6G ecosystem.