# 🔑 ****Federated Intrusion Detection (Privacy-Preserving)****

## **Research Problem**

Traditional intrusion detection systems (IDS) rely on centralized training using sensitive network traffic data. In real-world settings, organizations are unwilling (or legally restricted) from sharing raw logs across companies or even departments. **Federated Learning (FL)** enables collaborative IDS training without exposing raw data, but:

* Standard aggregation methods (FedAvg, FedProx) degrade under **non-IID client data** (different traffic distributions per organization).
* Current work lacks **cost-aware optimization** to balance detection accuracy, fairness across clients, and communication efficiency.

**Open problem:**  
👉 How can we design **optimization-driven federated aggregation strategies** that improve IDS performance on non-IID data, while being lightweight enough for deployment?

## **Datasets / Environments**

* **NSL-KDD** (link): Classic IDS dataset with normal vs multiple attack categories.
* **CICIDS2017** (link): Realistic modern intrusion dataset with DoS, brute force, infiltration, botnet traffic.
* **Simulation setup:** Split dataset into multiple "clients" with non-IID partitions (e.g., one client sees mostly DoS, another sees brute force).

## **Baselines**

1. **Centralized Training**: CNN/MLP model trained on pooled dataset.
2. **FedAvg** (McMahan et al., 2017): Classic averaging of client updates.
3. **FedProx** (Li et al., 2020): Improves FedAvg for heterogeneous data.
4. Optional: Simple ML models (Random Forest, XGBoost) trained locally on each client for comparison.

## **Your Contribution Angle**

1. **OOP Framework Design**
   * FederatedServer class: manages global model, aggregation.
   * Client class: trains locally, sends updates.
   * Aggregator module: handles aggregation strategy (pluggable for FedAvg, FedProx, Optimized).
2. **Optimization-Based Aggregation**
   * Implement **weighted aggregation** where weights are tuned by an **evolutionary algorithm (GA/PSO/DE)** instead of fixed by dataset size.
   * Objective: maximize **global F1-score**, minimize **variance across clients**.
3. **Evaluation**
   * Compare your optimized aggregation vs FedAvg/FedProx under:
     + IID vs Non-IID splits.
     + Different number of clients.
   * Metrics: Accuracy, Precision, Recall, F1-score, Communication cost.
4. **Expected Contribution**
   * A modular, OOP FL framework for IDS.
   * A novel **optimization-based federated aggregation method** that improves robustness on non-IID data.
   * Practical applicability for companies: better IDS collaboration without data sharing.