**Intrusion Detection System with  
 Federated-Learning**

The Schema:

1. Dataset Selection: three ids datasets

2. Parsing & Cleaning: remove the redundant and corrupted data, drop the meta data columns  
we need the network traffic system features that could be standraized

3. Missing values: feature-wise mean imputation

4. Normalization & Encoding:  
Numerical -> MinMax  
categorical -> OHE

5. Label Harmonization: remap attack labels into broader unified categories (**Benign, DoS/DDoS**, **Probe/Scan**, **R2L/U2R** (remote-to-local, user-to-root), **Other IoT Attacks** (from TON\_IoT telemetry/log anomalies)

-> to reduce class imbalance and create taxonomy

6. Design the schema (the paper didn’t give details so still exploring)

The FL Pipeline:

Models and parameters:

| **Method** | **Dataset** | **Model** | **Accuracy (%)** | **Learning Rate** | **Batch Size** | **Local Epochs** |
| --- | --- | --- | --- | --- | --- | --- |
| **FedAvg** | CICIDS2017 | CNN | 92.3 | 0.01 | 32 | 5 |
| **FedProx** | TON\_IoT | LSTM | 89.7 | 0.005 | 64 | 10 |
| **MOFL** | CICIDS2017 | Transformer | 93.5 | 0.001 | 32 | 5 |
| **Ours** | CICIDS2017 + TON\_IoT | Hybrid CNN–LSTM | 95.2 | 0.001 | 64 | 5 |