Distributed Anomaly Detection in Smart Grids Using Federated Learning

Abderrahmane Medabdellahi [C16023]
Master 2 – Artificial Intelligence & Data Science
University of Nouakchott
Supervisor: Dr. El Benany Med Mahmoud

February 10, 2026

Abstract

The increasing reliance on electrical power systems requires efficient monitoring mechanisms capable of detecting anomalies in real time. In distributed environments, traditional centralized solutions face limitations related to data privacy, communication overhead, and latency.

This project proposes a distributed anomaly detection framework based on a hierarchical Edge–Fog–Cloud architecture combined with Federated Learning. Local AutoEncoder models are trained at the Edge level using electrical measurements, while raw data remains locally confined. The Fog layer ensures real-time data streaming and coordination using Apache Kafka and Spark Streaming. Global model aggregation is performed in the Cloud using the Federated Averaging (FedAvg) algorithm.

The proposed approach enables collaborative learning without data centralization, improves scalability, and preserves data privacy, making it suitable for smart grid monitoring.

Keywords: Smart Grid, Anomaly Detection, Federated Learning, Edge Computing, Fog Computing

1. Introduction

Electrical power grids are complex infrastructures that require continuous supervision to ensure operational stability and service continuity. Abnormal conditions such as voltage fluctuations, overloads, or unexpected consumption patterns may lead to failures if not detected in a timely manner.

Conventional anomaly detection systems often rely on centralized architectures, where all measurement data is transmitted to a central server for analysis. Although effective, this paradigm introduces several challenges, including increased communication costs, latency, and concerns related to data confidentiality.

To overcome these limitations, this project adopts a decentralized approach based on Edge Computing and Federated Learning. By distributing the learning process across multiple local nodes and aggregating knowledge at a higher level, the proposed system aims to provide an efficient and privacy-preserving solution for anomaly detection in smart grids.

2. System Architecture

The overall architecture of the proposed system is illustrated in Figure 1. It follows a three-layer hierarchical model composed of Edge, Fog, and Cloud layers.

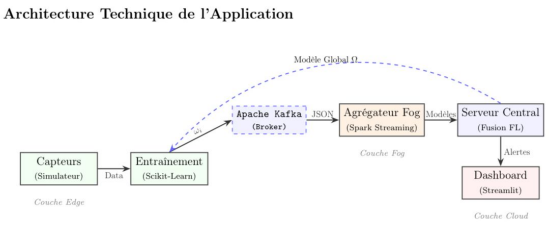

Figure 1: Edge–Fog–Cloud architecture for distributed anomaly detection based on Federated Learning

At the Edge layer, electrical measurements are collected and processed locally. Each Edge node trains a local AutoEncoder model using its own data, ensuring that raw measurements never leave the local environment. This design choice directly contributes to preserving data privacy.

The Fog layer acts as an intermediate coordination layer. It relies on Apache Kafka to manage real-time data streams and on Spark Streaming to perform intermediate aggregation and filtering. This layer reduces communication overhead and improves system responsiveness.

Finally, the Cloud layer is responsible for the global aggregation of local model parameters using the Federated Averaging algorithm. The resulting global model is redistributed to the Edge nodes, enabling continuous improvement of local anomaly detection performance.

3. Methodology

The anomaly detection process is based on an unsupervised learning approach using AutoEncoder models. Each local model is trained to reconstruct normal electrical behavior observed in historical data. During operation, an anomaly is detected when the reconstruction error exceeds a predefined threshold, indicating a deviation from normal behavior.

Instead of sharing raw measurement data, each Edge

node transmits only its learned model parameters to the Cloud. The global model is updated according to the Federated Averaging strategy:

$$\omega_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^{k} \tag{1}$$

where n_k denotes the number of samples used for local training at node k, and n represents the total number of samples across all nodes.

4. Experimental Setup

The proposed system was implemented using containerized services to emulate a distributed environment.

- **Dataset:** Electrical power consumption data partitioned across multiple Edge nodes
- **Model:** Neural AutoEncoder implemented using scikit-learn (MLPRegressor)
- **Infrastructure:** Multiple Edge nodes, one Fog layer (Kafka and Spark), and one Cloud server
- **Tools:** Python, scikit-learn, Apache Kafka, Spark Streaming, Docker

5. Results and Discussion

Experimental observations indicate that local AutoEncoder models are able to capture normal operating behavior and detect anomalous patterns effectively. The global model converges after several aggregation rounds, demonstrating that collaborative learning can be achieved without centralizing sensitive data.

The separation between data processing, coordination, and global learning, as shown in Figure 1, contributes to system scalability and robustness. The Fog layer plays a key role in reducing latency and handling real-time data flows.

6. Conclusion

This project presents a distributed anomaly detection framework for smart grids based on Federated Learning and an Edge–Fog–Cloud architecture. The proposed approach demonstrates that effective anomaly detection can be achieved while preserving data privacy and reducing communication overhead.

Future work will focus on improving model interpretability, integrating real-world sensor data, and evaluating the system under large-scale deployment scenarios.