# Development of Unsupervised Learning based Model for Cyber Attack Detection in Power Grid

Anushtha Tamrakar, Kunal Ajay Wasnik, Syed Lateefuddin, Punit Rathore

## Objective

- To develop an unsupervised approach that is computationally fast and accurate, has high degree of specificity and sensitivity, and cost-effective.
- Focus on minimizing false alarm rates, ensuring adaptability to dynamic environments.
- A GPU-integrated machine learning-based intrusion detection system prototype that is quick, precise, and economical by implementing a pilot experiment to validate the effectiveness and scalability of the proposed intrusion detection solution on IISc testbeds.

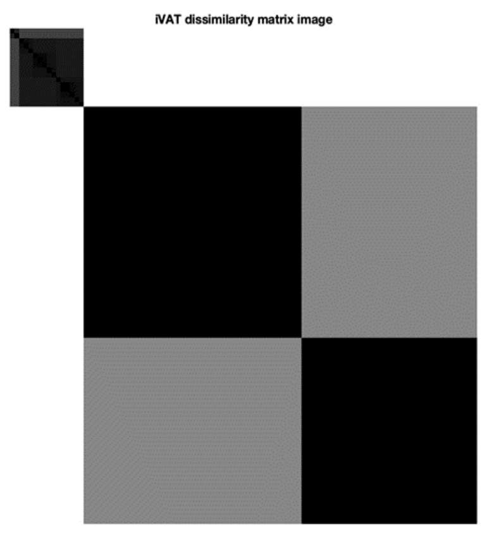
## Project Details

- The Indian Power Grid uses IEC 61850 and IEC 60870-5-104 to communicate between the various substation.
- IEC 61850 architecture consists of three layers, IEC 61850 consists of three layers: the station, process, and bay levels, communicating through the MMS, GOOSE, and SMV protocols.
- We utilize the publicly available datasets like IEC61850SecurityDataset and ICS Smart Grid Dataset for the IEC 60870-5-104.
- The proposed model developing to detect known attacks such as Malformed Packet Attacks, DoS, Address Resolution Protocol (ARP) Spoofing Attacks and man-in-the-middle (MITM) attacks.
- Once the models have been developed and evaluated, they will be deployed over NVIDIA Jetson computing units, and a pilot study will be conducted to evaluate the performance of the models via simulations.

**Progress until last review (April-May'24):**

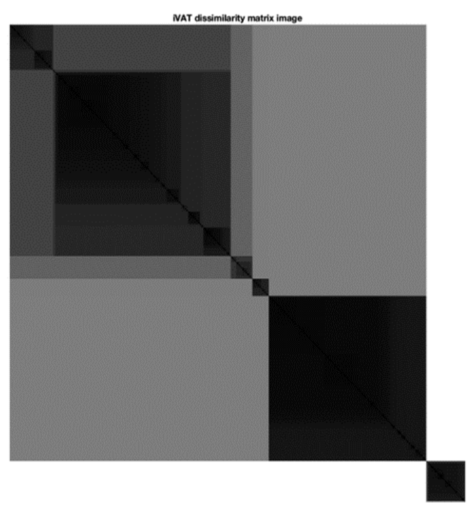
- Various supervised models have been tested to check the predictability of the features
- Various clustering-based unsupervised algorithms have been applied, and the results are as follows:

| | K-means | FCM | DBSCAN | linkage | VAT |
|---|---|---|---|---|---|
| Overall clustering accuracy | 99.98 | 50.05 | 84.76 | 42.56 | 99.91 |
| Attack accuracy | 99.98 | 48.33 | 100 | 50.05 | 99.90 |

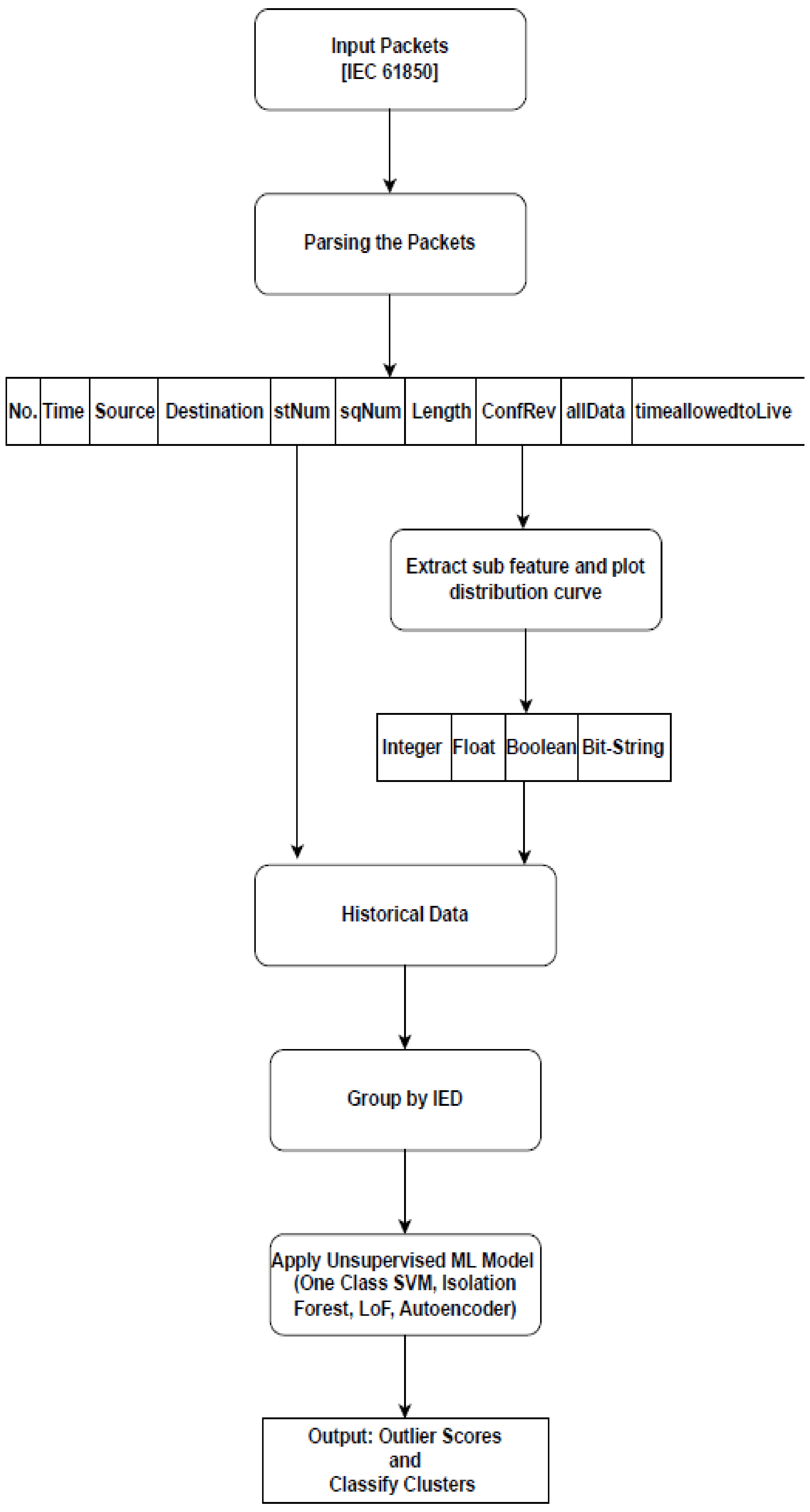


**IEC 61850 Dataset**

| | K-means | FCM | DBSCAN | Single Linkage | VAT |
|---|---|---|---|---|---|
| Overall accuracy | 72.8 | 65.7 | 61.3 | 63.5 | 80.2 |
| Attack accuracy | 70.1 | 64.3 | 63.7 | 62.4 | 90.2 |



**IEC 60870-5-104 Dataset**



Input Packets [IEC 61850]

Parsing the Packets

| No. | Time | Source | Destination | stNum | sqNum | Length | ConfRev | allData | timeallowedtoLive |
|---|---|---|---|---|---|---|---|---|---|

Extract sub feature and plot distribution curve

| Integer | Float | Boolean | Bit-String |
|---|---|---|---|

Historical Data

Group by IED

Apply Unsupervised ML Model (One Class SVM, Isolation Forest, LoF, Autoencoder)

Output: Outlier Scores and Classify Clusters

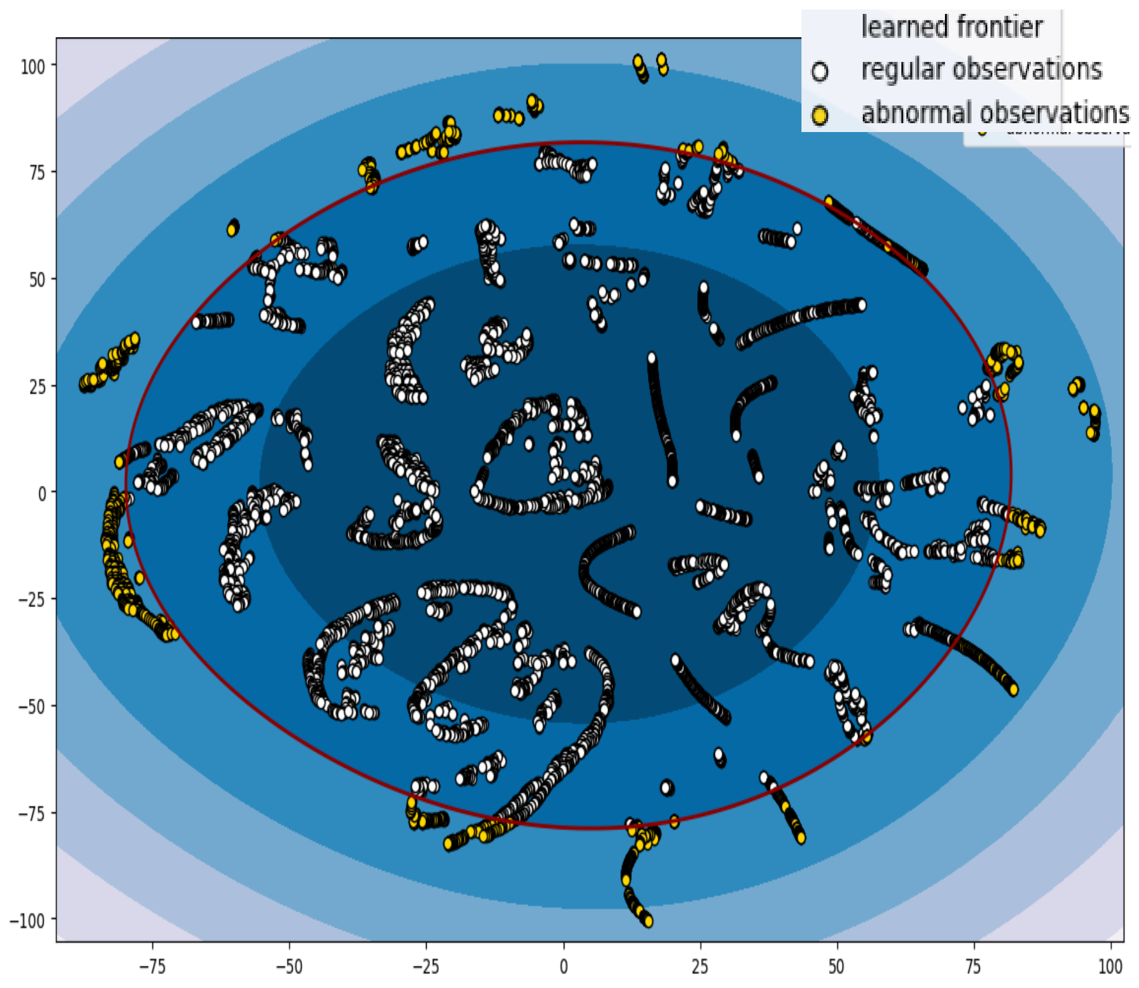
**Flow Diagram**

## Current Status

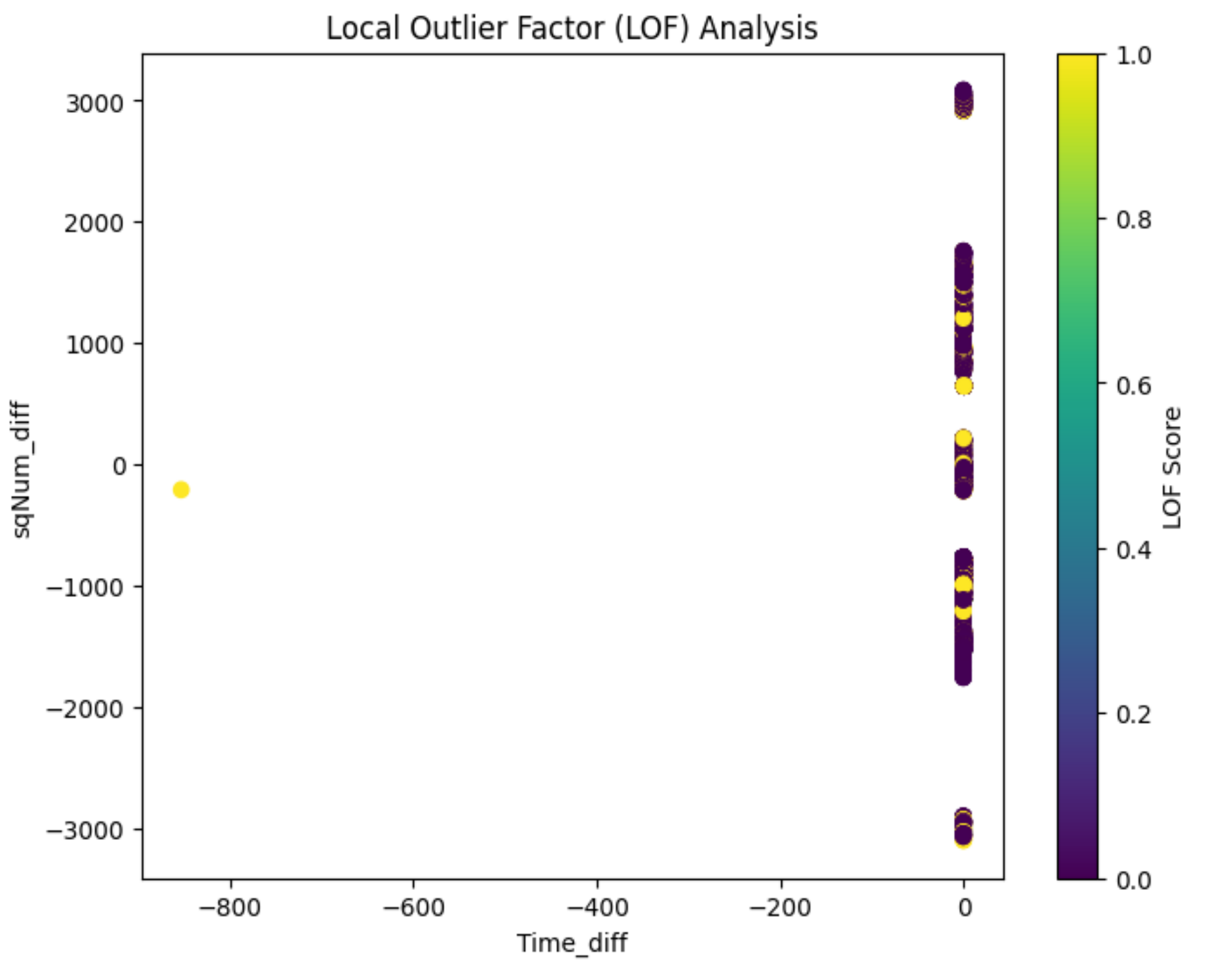
- Implemented Unsupervised-learning based anomaly detection algorithms viz., Isolation Forest, One class SVM (OCSVM) and Local Outlier Factor (LoF).
- Tested these models on the IEC 60870-5-104 datasets and other available GOOSE protocol Datasets
- **Ongoing**: (i) Testing and Validation on other datasets (ii) New features extraction (iii) Exploration of alternative unsupervised-learning methods

| Algorithms | Overall Accuracy | Precision (0 \| 1) | Recall (0 \| 1) | F1-Score (0 \| 1) |
|---|---|---|---|---|
| Isolation Forest | 0.79 | 0.80 \| 0.66 | 0.96 \| 0.27 | 0.87 \| 0.38 |
| OCSVM | 0.72 | 0.76 \| 0.29 | 0.91 \| 0.12 | 0.83 \| 0.17 |
| LoF | 0.66 | 0.75 \| 0.21 | 0.82 \| 0.15 | 0.78 \| 0.17 |

{ 0 | 1 } : 0 - Normal data Points 1 - Abnormal data points



OCSVM



LoF

POWERGRID CENTER OF EXCELLENCE IN CYBER SECURITY

RBCCPS Robert Bosch Centre for Cyber-Physical Systems