

Understanding Cyber Physical Events in IEC 61850 Digital Substations

Supervision: Julian Oelhaf (Computer Science 5) and Mamdouh Muhammad (Computer Science 7).

Motivation

A digital substation connects electrical measurements, IEC 61850 communication and protection decisions. Physical faults and cyberattacks may produce overlapping observations, but their initiating mechanisms and event sequences can differ. The dataset introduced by Cibin et al. [1] records several layers of this process. Understanding its labels, timing and feature relationships is necessary before drawing conclusions from a classifier.

Problem Statement

It is unclear which recorded features reliably distinguish physical faults from cyberattacks and which apparent differences reflect scenario design, label timing or acquisition artifacts. The thesis will establish what information each layer contains, whether engineered cross-layer features add distinct information, and whether these relationships remain useful on independent experimental runs.

Objective

The thesis will characterize the dataset and event mechanisms, analyze individual features and feature groups, and test physically and protocol-motivated feature engineering. Its ML contribution is a controlled comparison of raw, engineered, single-layer and combined cross-layer representations. Interpretable models will quantify accessible information and expose errors and shortcuts. The goal is an evidence-based account of the task, not a maximum accuracy score.

Dataset

The starting point is the public Cibin et al. digital-substation dataset [1], described as containing RTDS electrical measurements, IEC 61850 GOOSE and Sampled Values traffic, protection-related observations and virtual IED metrics across normal, fault and attack scenarios. The student will verify the released fields, scenario taxonomy, time alignment, label provenance and independent unit of analysis. Fault-versus-attack attribution or early detection will be defined only where the audited labels and timestamps support it.

Methodology

- Audit files, labels, missingness and synchronization; reconstruct representative event timelines and identify run, scenario and acquisition dependencies.
- Study distributions, temporal changes, feature-event associations and redundancy within and across physical, communication, protection and system feature groups.
- Implement documented electrical, protocol and cross-layer features with explicit formulas, observation windows and checks that each feature would be available at the stated decision time.
- Build a reproducible ML pipeline with a trivial baseline, regularized linear models and a modest non-linear model. Compare raw versus engineered features, individual groups, each modality and combined layers under identical splits and tuning budgets.
- Keep all observations from each experimental run together across training and evaluation. Report macro-F1, per-class recall and confusion matrices for supported classification tasks; analyze errors, modality and feature-group ablations, and held-out scenarios where sample coverage permits.

Expected Outcomes

- A data dictionary, event taxonomy, aligned example timelines and a record of label and timing uncertainty.
- An interpretable feature catalog showing which observations are informative, redundant or potentially misleading.
- A reproducible extraction and ML evaluation pipeline with grouped splits, controlled representation comparisons and error analysis.
- A clear account of what the dataset can support and a prioritized basis for follow-up cyber-physical ML research.

Indicative Timeline

- Months 1 - 2: Literature and dataset audit; clarify labels and timing, reconstruct events and agree on tasks and split rules.
- Months 3 - 4: Feature analysis and engineering; implement the extraction pipeline and initial ML baselines.
- Months 5 - 6: Controlled comparisons, ablations and error analysis; write the thesis and prepare reproducible handover artifacts.

Candidate and Supervision

The thesis suits a master's student with strong Python skills and a foundation in ML; FAU's Deep Learning lecture or comparable coursework is expected. Interest in power systems, cybersecurity or network protocols is useful. Weekly project meetings take place in person in Erlangen on Mondays at 14:00 - 15:30. PRL thesis students also participate in the relevant colloquium and give introduction and final presentations.

[1] N. Cibirin et al., "Cyber-Physical Power System Dataset for Cyber Security of Digital Substations," IEEE Access, 2026.
DOI: 10.1109/ACCESS.2026.3699153

[2] Sechi, F., Noori, N.S., Sharma, C. et al. Towards distinguishing cybersecurity attacks and safety faults in distributed energy resources-rich smart grids: a systematic literature review. Int. J. Inf. Secur. 25, 160 (2026).
<https://doi.org/10.1007/s10207-026-01308-z>