Generating Synthetic Substation Communication Profiles for Cross-Substation Transfer Learning

Filip Natvig
Dept. of Electrical Engineering
Uppsala University
Uppsala, Sweden
filip.natvig@angstrom.uu.se

Lars Nordström
Div. of Electric Power and Energy Systems
KTH – Royal institute of technology
Stockholm, Sweden
larsno@kth.se

Göran N. Ericsson
Dept. of Electrical Engineering
Uppsala University
Uppsala, Sweden
goran.n.ericsson@uu.se

The increasing digitalization of power systems introduces new cybersecurity risks, particularly in IEC 61850-based substations where the availability of communication services is crucial to operational safety. Anomaly-based intrusion detection systems (IDSs) are a promising countermeasure, which are particularly useful for detecting threats such as Denial of Service (DoS) attacks. However, developing and evaluating IDSs in this domain remains a challenge due to the scarcity of representative, labeled datasets and the difficulty of collecting traffic from multiple substations with different architectures.

To address these limitations, recent research has explored cross-substation transfer learning (XSTL)—a technique where an IDS model trained on data from one substation (the source domain) is reused to enhance detection performance in another (the target domain). This approach has the potential to reduce reliance on substation-specific training data and improve scalability. However, studying and validating XSTL requires datasets from several different substations, which are often unavailable due to cost, complexity, or proprietary restrictions.

In response, we propose a dataset manipulation technique for creating synthetic communication profiles from a single comprehensive dataset. By segmenting and reorganizing existing network traces, we simulate substations with varying traffic volumes, message patterns, and activity levels. Each resulting sub-dataset reflects a distinct communication profile that can be used in place of real data from multiple substations.

We validate this technique by statistically comparing the manipulated profiles against a reference dataset originating from a separate substation. The results show that the synthetic profiles capture meaningful variation in communication behavior, supporting their potential use in future XSTL experiments.

By enabling the creation of multiple synthetic domains from a single dataset, this technique provides a practical way to support the systematic evaluation of XSTL. In particular, it allows for controlled experimentation with different choices of source substation data, helping researchers investigate how communication profile characteristics influence transferability. This opens up new opportunities for understanding and improving the conditions under which XSTL is most effective in IEC 61850-based cybersecurity applications.