Towards Real-Time Distinction of Power System Faults and Cyber Attacks

Ali Abedi, Vetrivel S. Rajkumar, Alexandru Ştefanov, Peter Palensky

Department of Electrical Sustainable Energy
Delft University of Technology
Delft, The Netherlands
a.abedi-1@tudelft.nl

***Abstract*—This paper presents a methodology to distinguish between three-phase faults and GOOSE cyber attacks, aimed at opening the circuit breakers in the power grid. We propose a scheme that utilizes Phasor Measurement Unit (PMU)-enabled monitoring of power grid states, and communication network packet logs in the substation. In this scheme, by leveraging both cyber and physical data correlations and applying a Seasonal Autoregressive Moving Average (SARMA) model, we successfully distinguish between 3-phase faults and cyber attacks. The proposed scheme is tested using the benchmark IEEE 9-bus system, and can distinguish cyber attacks from faults in less than 0.2s. This demonstrates the usefulness of the proposed scheme for power system cyber security analytics.**

***Keywords*—anomaly detection, cyber attacks, cyber security, power system faults, synchrophasor**

I. Introduction

The electrical power system is undergoing a paradigm shift due to the energy transition and digitalisation. Thereby, it is being transformed into a complex Cyber-Physical System (CPS). This CPS offers numerous advantages, but also introduces cyber security issues. This is a real-world threat, as demonstrated by the cyber attacks on the power grid in Ukraine in 2015 and 2016 that led to power outages [1]. Hence, cyber security of power grids is of paramount importance. Under a faulted condition in the power system, protection relays detect and clear the fault by opening the associated circuit breakers. However, the communication infrastructure inside a substation can be exploited to force a malicious trip or opening of circuit breakers, even in the absence of any electrical fault. Such an attack threat is already shown by the aforementioned cyber attacks in Ukraine. Distinguishing these two situations is a challenging task due to the multi-domain nature of the problem, involving both cyber and physical elements.

Electrical fault signatures are well-studied and can be effectively captured by physical models of the power system. However, detection of stealthy cyber attacks within substations is non-trivial if purely based on physical power system measurements. Likewise, as noted in [2], detection or fingerprinting of cyber attacks at the bay level is a challenge. This is due to the differences between substation Operational Technology (OT) network traffic volumes at the bay and station level.

In related work, general model-driven and data-driven anomaly detection techniques that use measurements to detect anomalous behaviour in the power system operation are discussed in [3]. Also, intrusion detection methods that utilize communication packets to detect anomalous behaviour are presented in [4]. However, these works solely focus on the physical and cyber aspects, respectively. In [5] and [6], a detailed comparison is performed for power system disturbance and cyber attack discrimination using Machine Learning (ML). The authors consider PMU measurements and use network logs collected through SNORT and Syslog. Hence, actual network traffic flows are not considered. Furthermore, as also concluded by the authors, these supervised ML techniques require large datasets with labels, making them difficult to deploy in the control centre. In [2], the authors studied three types of cyber attacks including Denial-of-Service (DoS) and Manufacturing Messaging Specification (MMS) in a digital substation and investigate packet modification. However, they inspect features of individual packets, resulting in a significant overhead in the detection process. Hence, they are unsuitable for near real-time detection. In [7], PMU measurements are used to detect bad or missing data using LSTM. A deep auto-encoder and ridge classifier are trained based on PMU and captured MMS packets, respectively in order to diagnose the root cause of the failure. However, this work only considers component failures inside a substation and not actual electrical faults, i.e., short-circuits. Hence, correlations between cyber-physical system data for distinction between power system faults and cyber attacks is an open problem that our research seeks to address.

In this paper, we propose a data-driven scheme to distinguish between normal operations, three-phase faults, and stealthy cyber attacks targeting IEC 61850 in digital substations. This scheme can be implemented in a utility control centre for fast detection and distinction between these system modes. The scientific contributions of this work are as follows:

1. Formulating a data-driven architecture for detection and distinction between three-phase faults and cyber attacks in digital substations.
2. Application of the proposed architecture for situational awareness in cyber-physical power systems.

Following the intuition in this paper, one can devise similar schemes for distinguishing different types of faults and cyber attacks on power systems.

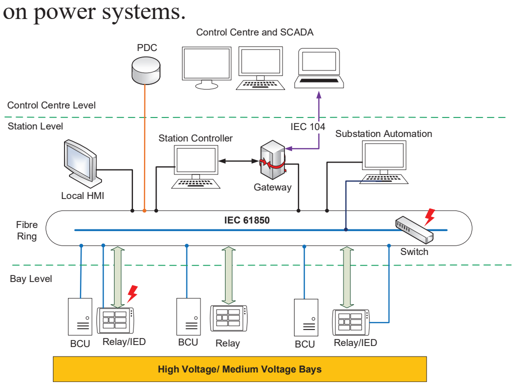

Fig. 1. Digital substation architecture. The cyber attacks in this paper focus on the bay and station level, indicated by the red symbols.

II. Problem Statement

Figure 1 depicts the architecture of a typical digital substation, based on the IEC 61850 standard considered in this paper. In the following sections, we briefly describe the type of cyber attack investigated in this paper and formally define the distinction problem, respectively.

A. IEC 61850 Cyber Attacks

Digital substations utilize IEC 61850 standard for communicating measurements and commands in the substation. Cyber security shortcomings in the standard, such as lack of encryption and authentication, makes digital substations vulnerable to cyber attacks. In this work, we consider the spoofing of IEC 61850 traffic, encapsulated as layer-2 Ethernet frames, within a digital substation. In the IEC 61850 standard, the Generic Object Oriented Substation Event (GOOSE) protocol is used to communicate control commands, such as trip or block, between Intelligent Electrical Devices (IEDs).

The GOOSE data payload contains the breaker statuses and circuit breaker controls. In the processing algorithm of GOOSE frames, the sequence number, i.e., sqNum, is continuously incremented with every sent GOOSE frame, while the status number is fixed. This holds true under normal operating conditions, wherein all GOOSE messages are communicated within a predefined time T ~ 100 to 5000 ms. In case of a substation event or fault, the status number, i.e., StNum, is changed by one and the sequence number is reset to zero. Furthermore, these event mode GOOSE frames are sent at a high rate of 0.5 to 5 ms. Consequently, by spoofing these event mode GOOSE frames, relays can be maliciously tripped, leading to unwanted opening of circuit breakers. This may result in disconnection of transmission lines or even generators.

The malicious spoofing does not cause changes to physical PMU measurements such as voltages and currents, unlike an electrical fault. This forms the rationale for the proposed distinction scheme of this paper. As illustrated in Figure 1, in the considered attack scenario, the attacker compromises the substation network switch, which interconnects the bay level devices including protection relays to substation level devices. Thus, the attacker is able to inject spoofed GOOSE messages, resulting in physical damages. A more detailed description of this type of cyber attack vector can be found in [8].

B. Distinction Problem

In the event of a transmission line fault, distance relays detect and trip to open circuit breakers to clear the fault. With communication-assisted protection, these relays also store and send the event details back to the control centre. However, as described in subsection II A, cyber attacks may cause the protection relay to trip in the absence of an actual fault event. This may result in misperception of the event as a fault in the control centre. In this study, we consider seven possibilities of a circuit breaker state: 1) it opens maliciously due to cyber attacks, as defined in II A; 2) it opens due to protection actions against faults; 3) it does not open due to the absence of fault; 4) it opens due to relay or circuit breaker malfunction when there is no fault; 5) it does not open due to a malfunction when there is a fault. 6) it opens due to remote operation by system operators via Supervisory Control and Data Acquisition (SCADA) system; 7) it does not open because of the malfunction of the SCADA system when the operator sends the open command.

Amongst these possibilities, we are interested in distinguishing between case 1 and case 2/3. Formally put, in a power system with *n* number of substations, in case of a circuit breaker opening in substation *i*, we aim to find out whether it is due to a regular fault or a cyber attack. Cases 4 and 5 are related to the reliability of protection equipment, outside the scope of this work. Also, cases 6 and 7 are excluded from this study since we focus on the automatic action of the relays, and not manual intervention by human operators. In the following subsections, we elucidate a few key assumptions considered in this paper, before proceeding to the proposed distinction approach.

C. Assumptions

First, we assume that the Phasor Measurement Unit (PMU) data collected from substations are not subject to cyber attacks, i.e., the integrity of PMU data is assured. Second, we assume that the GOOSE information is also periodically sent to the control centre SCADA system at a rate of 1-2 Hz using Routable-GOOSE (R-GOOSE), as described in [9]. Within the substation, network traffic data is logged, every 100 ms. Furthermore, the PMUs are optimally placed in terms of power system observability [10], such that the control centre has access to instant values of the phasors at each substation. Finally, it is important to assume that the power system does not destabilise, in case of a short-duration three-phase fault or in case of a line outage, e.g., *N-1* contingency.

III. Distinction Methodology

A. Proposed Distinction Scheme

As shown in Figure 2, the proposed detection and distinction module is implemented in the control centre. The inputs are the real-time PMU measurements from the Phasor Data Concentrator (PDC), logged GOOSE packet traffic as described above, and circuit breaker status at substation *i.* The GOOSE packet traffic rate reflects the inner dynamics and behaviour of the OT communication system at the bay level of substation *i*. Hence, its observation may reveal ongoing cyber events, i.e., provides cyber situational awareness. In case of a substation bay level event, i.e., relay trip or block, there should be strong correlation between the observed anomalies or deviations in the GOOSE network traffic and observed electrical fault signatures. In the absence of the latter, there is a high chance of a GOOSE spoofing cyber attack, as defined in Section II A. This is the rationale behind the proposed scheme depicted in Figure 3.

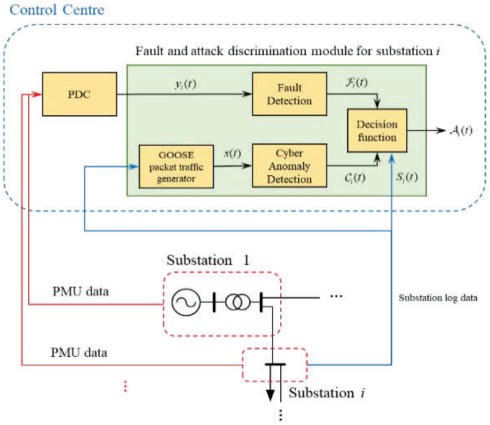

Fig. 2. Proposed fault and cyber attack distinction architecture.

The proposed data-driven distinction module is comprised of fault detection and cyber anomaly detection blocks, both of which produce binary outputs, fed to the decision making block. Our objective is to design the functions $\mathcal{F}(t)$ and $\mathcal{C}_i(t)$ for fault detection and cyber anomaly detection, respectively. The decision making function for substation *i* is defined as:

$$\mathcal{A}_i(t)=\begin{cases}1, & \text{if } \mathcal{F}(t)=0,\ \mathcal{C}_i(t)=1, \text{ and } S_i(t)=1\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

where $S_i(t)\in\{0,1\}$ is the circuit breaker state at substation *i* at time *t*.

1) Cyber Anomaly Detection using SARMA

OT communication traffic within a substation follows a less stochastic behaviour, in comparison to IT network traffic. This is due to the time-critical nature of the underlying physical processes. Therefore, such a pattern or behaviour can be learnt and then used in the cyber anomaly detection block from Figure 2. The Auto-Regressive Moving Average (ARMA) is one of the most widely used methods to linearly model and predict stationary univariate time-series [11]. Stationarity of a time-series refers to time-independence of the properties of such time-series, i.e., mean and variance [12]. If $\{x(k)\}, k\in\mathbb{Z}_+$ is a discrete-time stationary and univariate time-series, its ARMA model will be of the form:

$$\begin{aligned}\hat{x}(k) &= c+\theta_1 x(k-1)+\ldots+\theta_p x(k-p)\\ &\quad +\varepsilon(k)+c_1\varepsilon(k-1)+\ldots+c_m\varepsilon(k-m)\end{aligned} \tag{2}$$

in which c, θ_1 to θ_p and c_1 to c_m are parameters to be learnt, $\varepsilon(t)$ is a white noise signal, and $\hat{x}(t)$ is the prediction at time *t*, while *p* and *m* are the autoregressive and moving average orders, respectively. However, this model cannot capture the seasonality, i.e., a pattern that repeats over *n* time periods. This is clearly seen through Figure 3 that depicts the GOOSE packet traffic in this study under normal condition.

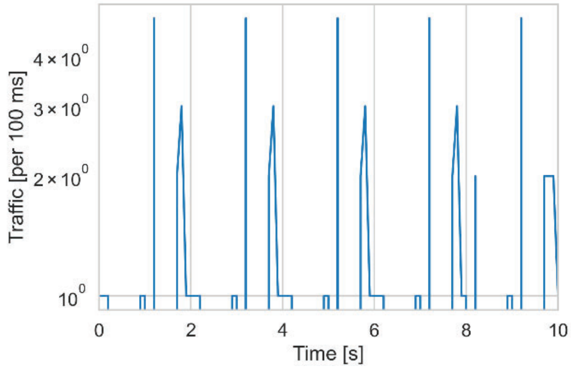

Fig. 3. Observed seasonality in GOOSE network traffic.

To overcome the non-stationarity imposed by the seasonality in the data, we use the Seasonal ARMA (SARMA). In SARMA, a seasonal difference $\alpha(x(t)-x(t-q))$ term is added to (1), where α is a linear coefficient to be learnt and q is the seasonal moving average order. The maximum likelihood via Kalman filter is used to fit the parameters of the SARMA model. Using the trained SARMA model, the function $\mathcal{C}_i(t)$ for anomaly detection is designed as the difference of the predicted value and real value at time *t*, with *T* being the threshold to be designed. This difference is called the residue signal.

$$\mathcal{C}_i(t)=\begin{cases}1, & \text{if } |\hat{x}(t)-x(t)|\geq T\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

2) Fault Detection

The three-phase fault is the most severe and commonly reported type of fault in transmission systems [13]. The post-fault voltage in case of a 3-phase or 3-phase to ground fault is computed as:

$$V_{pf}=\frac{Z_f V_f}{Z_1+Z_f} \tag{4}$$

where Z_f is the fault impedance, V_f is pre-fault voltage at the node/terminal and Z_1 is the positive-sequence impedance. In the above expression, Z_f is a variable, while, Z_1 and V_f are typically known to a utility. Therefore, we evaluate the maximum/minimum of V_f with regard to Z_f:

$$\frac{dV_{pf}}{dZ_f}=\frac{Z_f}{Z_1+Z_f} \tag{5}$$

Hence, the minimum and maximum post-fault voltages are functions of the fault and positive-sequence impedance. Typically in transmission systems $Z_f >> Z_1$, thereby, $V_{pf}\in\{0.5, 0.7\}$ p.u [13]. Therefore, we design the function $\mathcal{F}_i(t)$ for fault detection, as follows:

$$\mathcal{F}_i(t)=\begin{cases}1, & \text{if } T_1>|y_i(t)|>T_2\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

in which $y_i(t)$ is the bus voltage phasor measurement of substation i, with the thresholds $T_1=0.5$ and $T_2=0.7$.

IV. Case Study and Simulation Results

The benchmark IEEE 9-bus system is used to test the proposed distinction method between faults and spoofed GOOSE cyber attacks. A dataset is created consisting of one normal operation scenario, three fault scenarios, i.e., three-phase short-circuit, on different lines with different durations, and one spoofing cyber attack scenario of GOOSE data frames. This is achieved using a hardware-in-the-loop testbed consisting of Real-Time Digital Simulator (RTDS) interfaced with real IEDs and a network switch [10]. This is representative of the digital substation bay level. The fault locations are depicted in Figure 4. The cyber attack scenario is executed on substation 1 represented in Figure 4. The IED in this figure represents a distance protection relay. Also, we assume there are six PMUs, each of which is placed on buses 4-9 that measure voltage phasors in each phase.

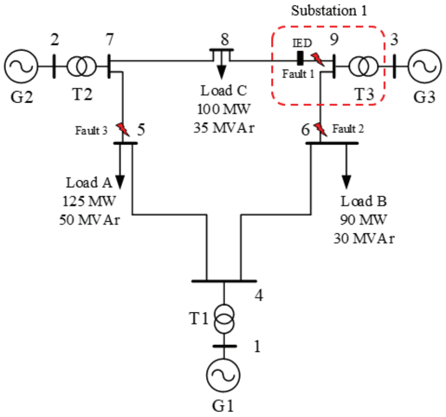

Fig. 4. IEEE 9-bus test system.

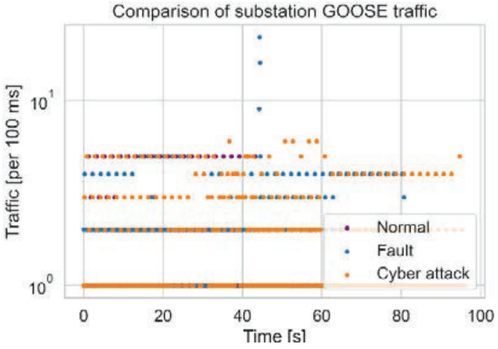

Fig. 5. Comparsion of GOOSE traffic under 3 conditions:.

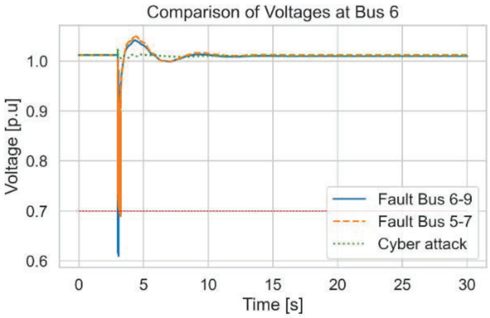

Fig. 6. Voltage phasors for three scenarios at bus 6.

Figure 5 depicts the GOOSE traffic under normal, fault, and cyber attack conditions. We observe that the normal and cyber attack cases are almost indistinguishable, rendering the detection of cyber attack a complicated task. Fig 6. shows the voltage magnitudes at bus 6 for the three scenarios, i.e., fault at bus 6-9, fault at bus 5-7, and a cyber attack at bus 6. It can be seen that in case of a fault in the power system, the bus voltage drops between 0.7 and 0.4 p.u. This is a similar case for others buses. The red dotted line depicts the upper threshold, i.e., 0.7 p.u. The correlations between GOOSE traffic and voltages for fault and cyber attack cases, in Figures 5 and 6 depict the distinction between a fault and cyber attack.

The autocorrelation of the GOOSE packet flow signal shows that the seasonality of this time series is 20 steps, i.e., it repeats every 2 seconds and can be observed in Figure 7. Hence, we set $q = 20$. This seasonality can also be seen from the periodic traffic pattern, as previously shown in Figure 3. Furthermore, by empirically setting $p = 5$ and $m = 0$, we obtain a good fit with the lowest mean square error. The SARMA model is trained with 250 out of 300 samples of the normal operation traffic data. The remaining 50 samples are subsequently used to test the trained model. The mean square error of this trained model is 0.36. Figure 8 illustrates the performance of the trained model on the 50 remaining test samples. The threshold T is empirically set to be 3.5 packets per 100 ms based on previous observations of the traffic rate in the normal case. Also, in this experiment $i = 1$.

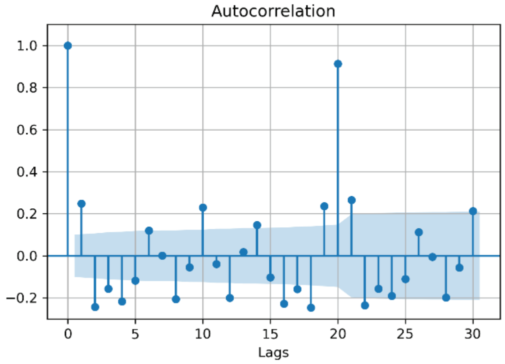

Fig. 7. Autocorrelation of GOOSE traffic time-series.

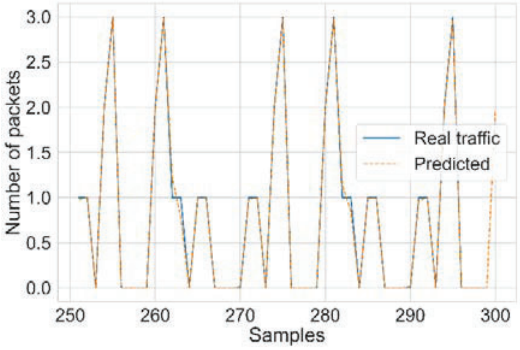

Fig. 8. Performance of the trained SARMA model.

Using the trained SARMA model, we apply the proposed scheme to the fault and attack scenarios. In Figures 9 and 10, in case of a fault occurrence and relay action, both the functions $\mathcal{C}_i(t) = 1$ and $\mathcal{F}_i(t) = 1$ in under a second. On the other hand, during the cyber attack, only the network traffic rate deviates from the prediction at some time instance, wherein $\mathcal{C}_i(t) = 1$ and $\mathcal{F}_i(t) = 0$.

In the cyber attack scenario, the attack starts from the beginning of the simulation, and after 8.47 seconds the circuit breaker connected to the IED in substation 1 opens. As depicted in Figure 11, this does not have a severe effect on the PMU measurements. However, immediately after this event, before the circuit breaker closes again, the cyber anomaly detection block detects an anomaly at around 8.6 seconds from the beginning of the simulation. Since $S_i(t) = 1$ and $\mathcal{F}_i(t) = 0$, a cyber attack alarm is generated by the decision making block. The overall time from circuit breaker opening until generating the cyber attack alarm is about 130 ms as marked in Figure 11. It is important to note, the performance of the proposed method is unaffected by the PMU communication latencies. These are typically in the order of 40-50 ms [14], sufficiently less than the moving window for detection used in our paper, i.e., ~1 s.

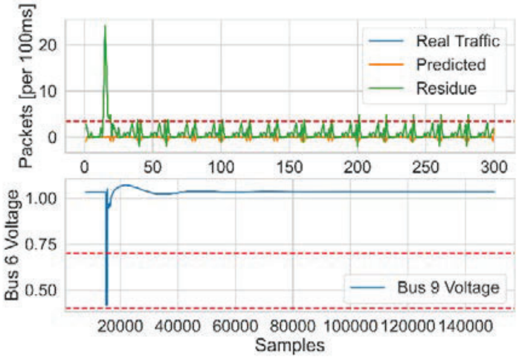

Fig. 9. Distinction results under the fault scenario 1. The red dashed lines indicate the thresholds.

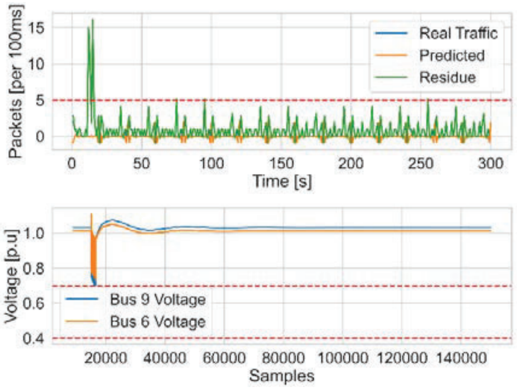

Fig. 10. Distinction results under the fault scenario 2. The red dashed lines indicate the thresholds.

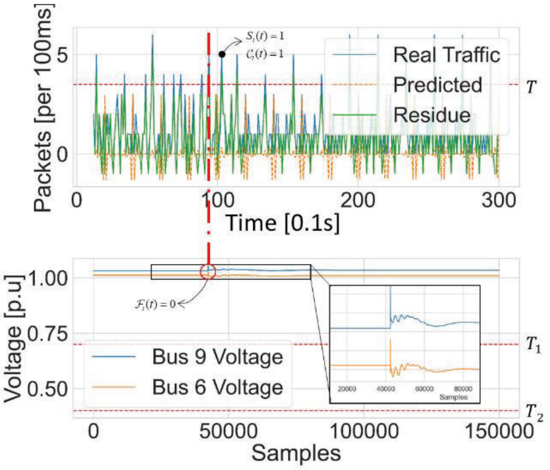

Fig. 11. Distinction results under the cyber attack scenario.

V. Conclusions and Future Work

In this paper, we proposed a scheme for real-time distinction between GOOSE cyber attacks in digital substations and three-phase electrical faults in the power grid. It utilizes the SARMA model and PMU measurements and GOOSE traffic in order to make a decision about the cause of a circuit breaker opening in a digital substation. From the experimental results, it is shown that this scheme can distinguish the cyber attacks and faults in less than 200 ms.

One limitation of the proposed scheme is that it needs an offline training stage based on normal operational GOOSE traffic. In future work, we will derive an online method to eliminate the offline learning. Furthermore, the existing methods such as the Receiver Operating Characteristic (ROC) curve, help find the optimal value for the threshold T to have a minimum false positive rate. This will be the focus of our future research.

VI. Acknowledgement

This work was partially supported by Designing Systems for Informed Resilience Engineering (DeSIRE) program of the 4TU Resilience Engineering Centre and EU H2020 projects: InnoCyPES (Grant Agreement Number 956433) and ERIGrid 2.0 (Grant Agreement Number 870620).

References

[1] L. Gjesvik and K. Szulecki, “Interpreting cyber-energy-security events: experts, social imaginaries, and policy discourses around the 2016 Ukraine blackout,” *Eur. Secur.*, vol. 0, no. 0, pp. 1–21, Jun. 2022.

[2] J. Parssinen, P. Raussi, S. Noponen, M. Opas, and J. Salonen, “The Digital Forensics of Cyber-Attacks at Electrical Power Grid Substation,” in *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*, Istanbul, Turkey, Jun. 2022, pp. 1–6.

[3] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, “A systems and control perspective of CPS security,” *Annu. Rev. Control*, vol. 47, pp. 394–411, Jun. 2019.

[4] Q. Liu, V. Hagenmeyer, and H. B. Keller, “A Review of Rule Learning-Based Intrusion Detection Systems and Their Prospects in Smart Grids,” *IEEE Access*, vol. 9, pp. 57542–57564, Apr. 2021.

[5] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, “Machine learning for power system disturbance and cyber-attack discrimination,” in *2014 7th International Symposium on Resilient Control Systems (ISRCS)*, Denver, CO, USA, Aug. 2014, pp. 1–8.

[6] G. Intriago and Y. Zhang, “Online Dictionary Learning Based Fault and Cyber Attack Detection for Power Systems,” in *2021 IEEE Power & Energy Society General Meeting (PESGM)*, Washington, DC, USA, Jul. 2021, pp. 1–5.

[7] A. Ahmed *et al.*, “Cyber Physical Security Analytics for Anomalies in Transmission Protection Systems,” *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 6313–6323, Nov. 2019.

[8] V. S. Rajkumar, M. Tealane, A. Ştefanov, A. Presekal, and P. Palensky, “Cyber Attacks on Power System Automation and Protection and Impact Analysis,” in *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, The Hague, Netherlands, Oct. 2020, pp. 247–254.

[9] M. Kanabar, A. Cioraca, and A. Johnson, “Wide Area Protection & Control using high-speed and secured Routable GOOSE Mechanism,” in *2016 69th Annual Conference for Protective Relay Engineers (CPRE)*, Apr. 2016, pp. 1–6.

[10] J. Qi, K. Sun, and W. Kang, “Optimal PMU Placement for Power System Dynamic State Estimation by Using Empirical Observability Gramian,” *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 2041–2054, Jul. 2015.

[11] O. Nelles, “Nonlinear Dynamic System Identification,” in *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*, O. Nelles, Ed. Berlin, Heidelberg: Springer, 2001, pp. 547–577.

[12] S. Makridakis, S. C. Wheelwright, and R. J. Hyndman, *Forecasting methods and applications*. John wiley & sons, 2008.

[13] M. H. J. Bollen, P. Goossens, and A. Robert, “Assessment of voltage dips in HV-networks: deduction of complex voltages from the measured RMS voltages,” *IEEE Trans. Power Deliv.*, vol. 19, no. 2, pp. 783–790, Apr. 2004.

[14] C. Lackner, F. Wilches-Bernal, B. J. Pierre, and D. A. Schoenwald, “A Tool to Characterize Delays and Packet Losses in Power Systems With Synchrophasor Data,” *IEEE Power Energy Technol. Syst. J.*, vol. 5, no. 4, pp. 117–128, Dec. 2018.