High-fidelity Intrusion Detection Datasets for Smart Grid Cybersecurity Research

Heng Chuan Tan*, Md Adeeb Hossain*, Daisuke Mashima*, Zbigniew Kalbarczyk†
* Illinois Advanced Research Center at Singapore (IARCS)
† University of Illinois at Urbana-Champaign
Email: {hc.tan, A.Hossain, Daisuke.m}@iarcs-create.edu.sg, kalbarcz@illinois.edu

***Abstract*—Intrusion Detection Systems (IDSes) are key defense mechanisms for securing smart grids against cyberattacks. They require realistic datasets to develop accurate models for detecting network anomalies. However, acquiring realistic datasets is challenging due to the need for expert knowledge to accurately label attack data, the sensitive nature of the information, and safety issues related to attacking the actual power systems. Consequently, there is a lack of high-fidelity datasets for testing and validating the efficacy of IDSes. While synthetic datasets provide a good workaround, they are often unrealistic and fail to capture the physical dynamics of power systems under cyberattacks. To address this gap, we leverage the Electric Power and Intelligent Control (EPIC) testbed, a hardware-based smart grid security testbed, to simulate false data injection attacks (FDIA) and time delay attacks (TDA) on two critical power grid operations, namely generator synchronization and reverse power prevention. Our goal is to generate representative datasets that accurately model those operations under normal and attack conditions. By making these datasets publicly available, we enable the research community to develop more effective IDS solutions to enhance the security of smart grids.**

***Index Terms*—Smart grids, Cybersecurity, Testbeds, False data injection attacks, Time delay attacks, Open-source Datasets**

I. Introduction

The advent of digital technologies has transformed the way smart grids operate and distribute electricity. A key transformation is the use of information and communication technology (ICT) to enable real-time monitoring and control of power devices. While this integration enhances operational efficiency, it also exposes the grid to cyberattacks that can disrupt power operations [1], [2]. To defend against these cyber threats, various intrusion detection systems (IDSes) utilizing statistical, rules-based, and machine-learning methods have been proposed for smart grids. [3]–[7].

These IDSes rely on realistic datasets to develop accurate models for anomaly detection. However, acquiring such datasets is challenging as it requires many man-hours and expertise to accurately label the data. Even if datasets are available, smart grid operators may be reluctant to share them for fear that they may expose additional vulnerabilities and increase the risk of exploitation. Due to the critical nature of smart grids and public safety concerns, it is also practically impossible to conduct attack experiments on real-world production systems to collect datasets.

The lack of high-fidelity datasets for IDS design has led researchers to rely on simulation models to generate synthetic datasets [8], [9]. However, simulations may not fully replicate the actual behavior of power systems and communication patterns. Thus, numerous smart grid testbeds have been established to allow researchers to simulate cyberattacks and contribute datasets in a safe and controlled environment [10], [11]. Despite these efforts, those public datasets are limited to cyberattacks that exploit industrial protocols (e.g., IEC 61850, DNP3, IEC 60870-5-101/104, Modbus TCP), with little consideration of their impacts on the grid operations.

This paper introduces a high-fidelity Electric Power and Intelligent Control (EPIC) testbed and discusses two fundamental operations commonly found in most power systems. Using ARP spoofing and man-in-the-middle (MITM) techniques, we launch false data injection attacks (FDIA) and time delay attacks (TDA) on actual power grid equipment in EPIC to disrupt generator synchronization and reverse power prevention operations. Specifically, FDIA aims to manipulate measurements and commands to mislead the control system into making wrong decisions, while TDA delays the transmission of control commands to affect the system state. Our main contribution is the generation of representative datasets that capture real-world attacks for validating IDS solutions. These open-source datasets[1] will enable researchers to benchmark IDS solutions and design specification rules and statistical models to detect cyberattacks that target critical grid operations. Additionally, researchers can perform time-series analysis using advanced machine learning algorithms such as recurrent neural networks to develop novel IDSes capable of identifying emerging threats.

The rest of the paper is organized as follows. Section II outlines the gaps in related works. Section III introduces the EPIC power grid testbed. Section IV details our attack implementation and scenarios, followed by dataset collection in Section V. Section VI concludes the paper.

II. Related Work

Biswas et al. proposed a framework for generating synthetic datasets for IEC 61850-based electrical substations [12]. Their framework utilizes substation configuration language (SCL) files obtained from intelligent electronic devices (IEDs) to model various substation scenarios that include normal, disturbance, and attack scenarios. For modeling the attack

[1]Available at https://github.com/smartgridadsc/EPIC_Attack_Datasets

scenarios, the authors focused primarily on denial-of-service (DoS) attacks and modification of the GOOSE payload to simulate message suppression and data modification attacks. In another work, the authors developed the framework into a toolchain that allows users to generate customized GOOSE datasets for cybersecurity research [13].

Quincozes et al. developed a framework that relies on realistic electrical measurements to support the generation of datasets. The electrical measurements are derived by modeling a real power grid using the Power Systems Computer Aided Design (PSCAD) simulation tool [9]. In modeling the attack scenarios, the authors considered five attack models (i.e., replay, masquerading, injection, high StNum, and DoS) and extracted 69 traffic features from the GOOSE and Sampled Values (SV) protocols to generate seven attack datasets. However, their attack scenarios are based on exploiting the inherent communication vulnerabilities in GOOSE and SV protocols. In contrast, we focus on commonly used power grid operations to design concrete attack scenarios to illustrate their disruptions to grid functionalities.

Radoglou-Grammatikis et al. published a DNP3 dataset containing nine cyberattacks, such as unauthorized commands, replay, and DoS attacks [14]. This dataset was generated using a testbed comprising multiple PCs that mimic the functions of Remote Terminal Units (RTUs), IEDs, and a human-machine interface (HMI). Three attacker PCs were introduced to the testbed to exploit various DNP3 function codes to send malicious commands, disrupt communication, and manipulate data. Although this synthetic dataset is useful for implementing machine learning-based IDSes, it does not consider the impact of cyberattacks on the physical processes.

Another direction involves developing smart grid testbeds for cybersecurity research and training. EPIC is one such testbed designed by the Singapore University of Design and Technology (SUTD) [11]. It has been utilized to support several hacking exercises, such as the Critical Infrastructure Security Showdown (CISS) [15] and the Critical Infrastructure Defence Exercise (CIDeX) [16]. Through these hacking exercises, only normal datasets that characterize the different operating scenarios of the EPIC testbed are shared with the research community [17]. Specifically, the datasets contain IEC 61850 Manufacturing Message Specification (MMS) messages and Modbus TCP packets for capturing automated protection and control within modernized substations. In this work, we expand these datasets by contributing detailed attack scenarios and corresponding attack datasets to guide IDS research.

The KASTEL Energy Lab at the Karlsruhe Institute of Technology (KIT) is a hybrid testbed that integrates physical components with software simulations [18]. This unique setup can simulate PV, wind turbines, battery systems, and transmission/distribution substations. The testbed is also compatible with many protocols, such as Modbus TCP, IEC 60870-5-104, Profinet, and IEC 61850. This rich protocol support allows for the design of experiments to investigate the interplay between various subsystems within the energy communication networks, providing insights into system behavior under different operating scenarios. However, as of this writing, no datasets have been made publicly available.

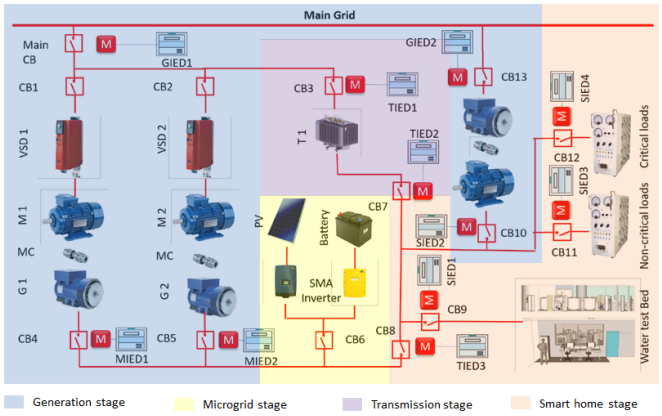

Fig. 1. EPIC architecture and its subcomponents [11]

III. The EPIC Testbed

A. Testbed Overview

EPIC is a modernized smart grid testbed that utilizes real-world power equipment to generate, transmit, and distribute electricity. As shown in Figure 1, the testbed consists of four stages: generation, transmission, microgrid, and smart home. These stages mimic real-world grid operations and are interconnected by power lines and communications buses.

The generation stage draws power from the mains and incorporates three generators rated at 10 kW each to simulate different generation profiles. The transmission stage contains transformers to step up or down voltage levels for distribution to the microgrid or smart home stage. The microgrid stage comprises distributed energy resources such as solar panels, energy storage systems, and configurable loads that can operate in grid-tied mode or islanded mode. Finally, the smart home stage replicates power consumption in residential and commercial settings. This stage consists of smart meters to measure and monitor energy consumption patterns.

Figure 2 depicts the network architecture, where each stage is monitored by a programmable logic controller (PLC) and a set of IEDs arranged in a ring network. The PLC uses the MMS protocol to communicate with the Supervisory Control and Data Acquisition (SCADA) workstation and IEDs. In addition, it uses the Modbus TCP protocol to interact with the actuators (i.e., variable speed drives — VSDs) to control the physical processes. The IEDs use the GOOSE protocol to control circuit breakers and exchange real-time information with adjacent IEDs to monitor the health of the grid. The MMS and GOOSE protocols are part of the IEC 61850 standard designed to improve device interoperability.

B. Testbed Operations

The smart home PLC (SPLC) plays a critical role in synchronizing new generators with the grid and preventing reverse power flows in the generators. These operations require

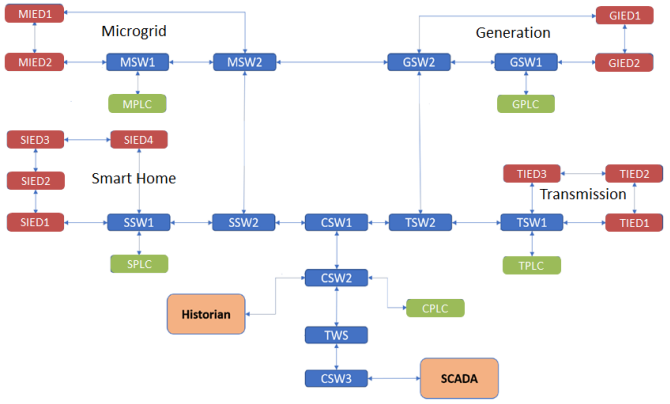

Fig. 2. The EPIC network diagram comprising Programmable Logic Controller (PLC), Intelligent Electronic Device (IED), and network switch (SW), where the prefix "G", "T", "M", and "S" represent the generation, transmission, microgrid and smart home stage, respectively. Additionally, TWS represents the firewall.

the SPLC to communicate with the SCADA and microgrid IEDs (MIEDs) to receive MMS commands. Upon receiving the commands, the SPLC issues Modbus commands to regulate the VSDs' speed associated with the generators.

Generator Synchronization: Synchronization is the process of connecting a generator to the grid. This is necessary when existing generators cannot meet the demand load or when there is a need to distribute the load among multiple power sources to improve grid stability. The purpose of synchronization is to ensure that the incoming generator's frequency matches that of the grid. Otherwise, an out-of-sync condition can damage the generator and cause disturbances to the power systems. This process is managed by SCADA, SPLC, and MIEDs, as shown in Figure 3. First, we assume G2 is connected to the grid. To synchronize G1, SCADA sends an MMS sync_start command to instruct the SPLC to accelerate VSD1 to 7502 (1500.4 rpm) while maintaining VSD2 at 7500 (1500 rpm) via Modbus. The difference in the two VSDs' speeds will cause the frequency of G1 to increase until it matches that of the reference generator G2. MIED2 monitors this frequency by measuring the phase angle difference between the two generators. When the phase angle difference approaches zero, MIED2 will send a GOOSE packet to close the circuit breaker (i.e., Q2C) and connect G1 in parallel with the reference generator (G2). Subsequently, MIED2 sends an MMS in_sync command to notify the SPLC that synchronization is complete. Finally, the SPLC issues a Modbus command to reset VSD1 back to its original speed of 7500 (1500 rpm).

Reverse Power Prevention: When the two generators are synchronized, they will rotate at nearly equal speeds to supply power to the grid. If one generator fails to maintain rotation due to faults, it will shift from generating mode to motoring. In motoring mode, the generator behaves like a motor and draws power from the grid, causing reverse power flows. Failure to prevent reverse power can damage the generator and cause it to trip. As shown in Figure 3, the MIEDs monitor reverse power

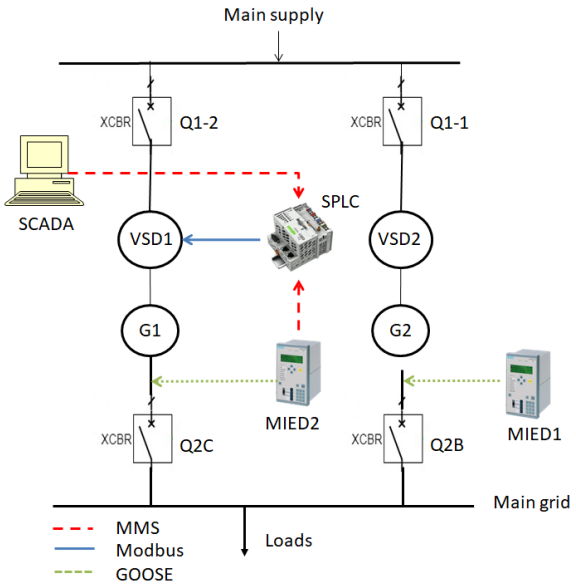

Fig. 3. Components and communication flows in the microgrid stage for synchronizing generators and monitoring reverse power flows in generators.

conditions in the generators. Upon detecting reverse power flows, the corresponding MIED sends an MMS command (i.e., gen1_p_negative == True or gen2_p_negative == True) to the SPLC. This command will instruct the SPLC to increase the VSD speed of the affected generator to 7501 (1500.2 rpm) while maintaining the other generator to operate at 7500 (1500 rpm) to counteract the impacts of reverse power.

IV. Attack Implementation

We exploit the lack of authentication in MMS and Modbus TCP protocols, which makes them susceptible to ARP spoofing and MITM attacks [19]. Additionally, we assume that the attackers are familiar with the power grid operations, allowing them to target specific packets and modify their payloads. Specifically, attackers can launch FDIA and TDA to disrupt the generator synchronization and reverse power prevention operations [20], [21]. In this work, FDIA refers to falsifying measurement data or commands into the network, while TDA involves delaying control commands from the SCADA or SPLC systems. Figure 4 shows the various attack entry points within EPIC.

A. Attack Tools and Methods

We download the dsniff package onto an attacker machine and use the ARP spoofing tool to intercept network packets between two target devices. Next, we develop a custom Scapy program to manipulate the intercepted MMS and/or Modbus packets. Specifically, our program uses nfqueue to route all intercepted MMS and Modbus packets to a queue in kernel space. This nfqueue allows Scapy to analyze and modify intercepted packets in real time before forwarding them to the intended destinations, thus preventing packet duplication that could disrupt the normal traffic flow. To redirect MMS and Modbus packets to the kernel queue, we configure iptables

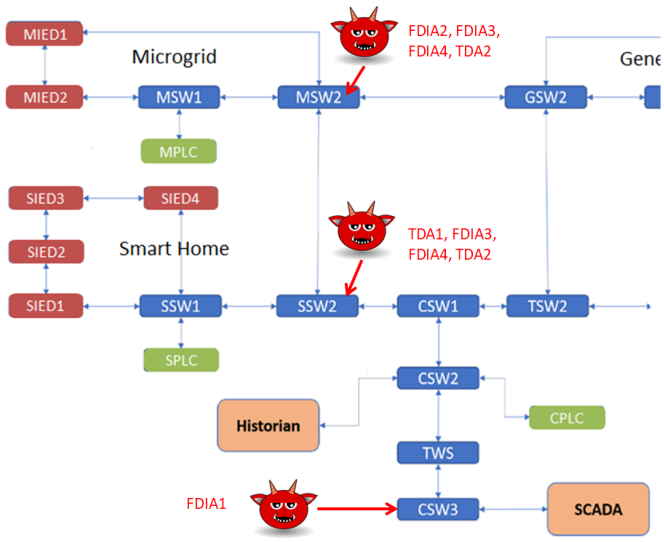

Fig. 4. The point of entries for various FDIA and TDA attacks

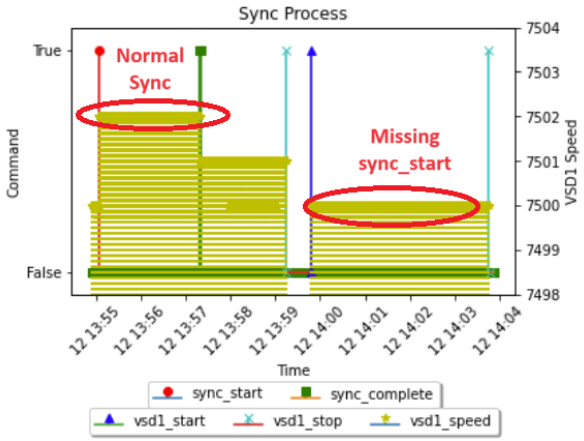

Fig. 5. FDIA1: sync_start commands are suppressed by the attacker to deny generator from synchronizing.

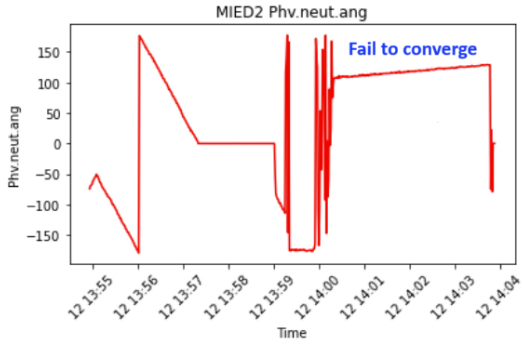

Fig. 6. FDIA1: The phase angle of G1 fails to converge.

rules with destination ports 102 and 502, respectively. Next, we implement filtering logic within our program to identify specific MMS packets for modification based on their MMS object reference (i.e., LD/LNFCDO$DA), where LD, LN, FC, DO, and DA represent a logical device, logical node, functional constraint, data object, and data attributes, as defined by the IEC 61850 standard [22]. Similarly, Modbus function codes are used to filter Modbus packets that require modification. These filtering rules ensure that only the correct MMS and Modbus packets are manipulated during the attack. We provide more details of each attack experiment in Table I and explain them further in subsequent sections.

B. Generator Synchronization Attacks

FDIA1: Deny G1 from synchronizing. As shown in Figure 4, we launch ARP spoofing at CSW3 to hijack MMS packets between the SCADA workstation and the SPLC. Upon detecting an MMS command from the SCADA to synchronize G1, we modify the sync command by setting its value to false. This attack prevents the SPLC from synchronizing G1 with the grid, resulting in imbalances in power flows and instability within the electric grid. As can be seen in Figure 5, normal synchronization is indicated by a circle marker (i.e., sync_start) and a square marker (i.e., sync_complete). The synchronization duration varies depending on the convergence speed of the phase angle to zero. Typically, it ranges from 1 to 2 minutes. Because sync_start is set to False, VSD1 fails to accelerate to 7502 (1500.4 rpm), causing the phase angle to never converge in Figure 6. Hence, G1 is unable to synchronize and remains isolated from the grid.

FDIA2: Incorrect synchronization status on SCADA. The SCADA receives phase angle updates from MIEDs via MMS packets. By modifying the phase angle value to a specific value at MSW2 in Figure 2, we trick the operator that the phase angle cannot converge and that synchronization is incomplete on the SCADA display. Due to this false information, operators are forced to take unnecessary steps to rectify the problem, which can lead to unexpected downtime and operational issues. Figure 7 and Figure 8 show how modifying phase angle data to a value of 30° can affect what an operator sees on the SCADA display.

TDA1: Prolong synchronization. Instead of accelerating VSD1 to achieve synchronization, we command the VSD1 to rotate at normal speed (i.e., 7500 or 1500 rpm), causing the synchronization process to take longer than expected. This delay can lead to an insufficient power supply, causing operational disruptions and a decrease in grid stability. This attack is achieved by spoofing Modbus packets between the SPLC and VSD1 at SSW2 in Figure 2. By continuously manipulating the VSD1 speed, we can prevent the generators from reaching the desired synchronized state. The generator G1 can only resume syncing once the attack is released, as illustrated by an increase in Modbus speed to 7502 and the presence of the sync_complete command in Figure 9.

C. Reverse Power Attacks

FDIA3: Cause motoring effect in G1. This attack aims to induce reverse power flows in G1 by ensuring that VSD2 rotates faster than VSD1. This attack is achieved by intercepting and modifying the Modbus commands sent by the SPLC to both VSDs at SSW2 in Figure 2. In this experiment, we set

TABLE I
DETAILS OF ATTACKS AND COMMANDS

Attack	Source	Destination	Type	Description	Remarks
FDIA1	SCADA	SPLC	MMS command	Modify sync_start from True to False	MMS object: GGIO17COSPCSO2$Oper
FDIA2	SPLC	MIED2	MMS measurement	Modify phase angle to any arbitrary value	MMS object: LLN0$Measurement
TDA1	SPLC	VSD1	Modbus command	Set Modbus speed to 7500 (or 1500 rpm)	N.A.
FDIA3	MIED2	SPLC	MMS command	Set gen1_p_negative to False	MMS object: GGIO1STInd5$stVal
	SPLC	VSD1	Modbus command	Set VSD1 to 7500 and VSD2 to 7501	N.A.
FDIA4	MIED2	SPLC	MMS command	Set gen1_p_negative to True	MMS object: GGIO1STInd5$stVal
	SPLC	VSD1	Modbus command	Set VSD1 speed <7501	N.A.
TDA2	MIED2	SPLC	MMS command	Set gen1_p_negative to True	MMS object: GGIO1STInd5$stVal
	SPLC	VSD1	Modbus command	Set VSD1 speed to 7500 to match VSD2	N.A.

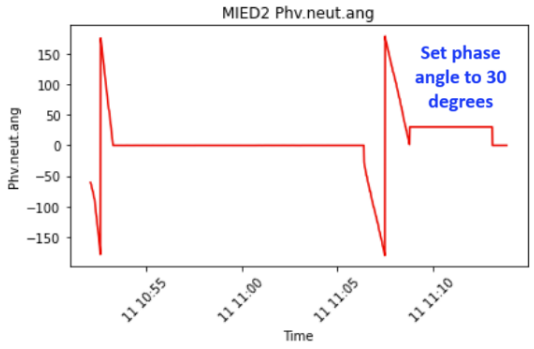

Fig. 7. FDIA2: An attacker modifies the phase angle to 30° to create a false representation of an incomplete generator synchronization.

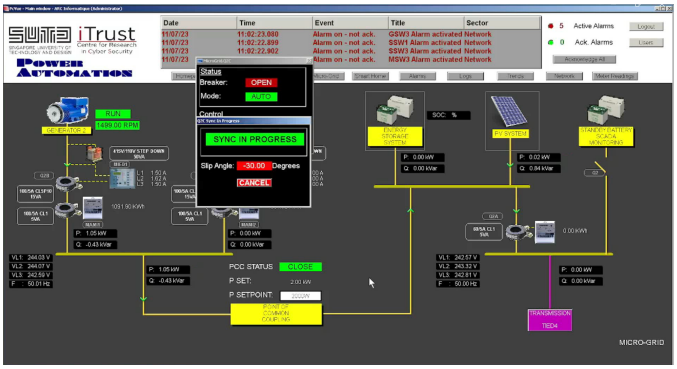

Fig. 8. Generator (G1) is synchronized with the grid. However, the phase angle shows 30° to indicate that sync is incomplete.

VSD2 to spin at 7501 (i.e., 1500.2 rpm) and VSD1 at 7500 (i.e., 1500 rpm). Since VSD2 speed is higher than VSD1, the SPLC cannot regulate the VSD1 speed when G1 enters reverse power mode. As a result, G1 is always drawing power from the grid, forcing it to go into motoring mode. Figure 10 illustrates how this attack decreases G1's total real power, which must be compensated by G2, causing it to operate under increased stress over a prolonged period and potentially leading to wear and tear of the equipment.

FDIA4: Cause G1 to trip. In the previous FDIA3 attack, we increased VSD2 speed slightly to cause G1 to enter reverse power mode. In FDIA4, we underspin VSD1, causing G1 to experience a substantial increase in reverse power. This attack involves manipulating two control variables. First, we initiate

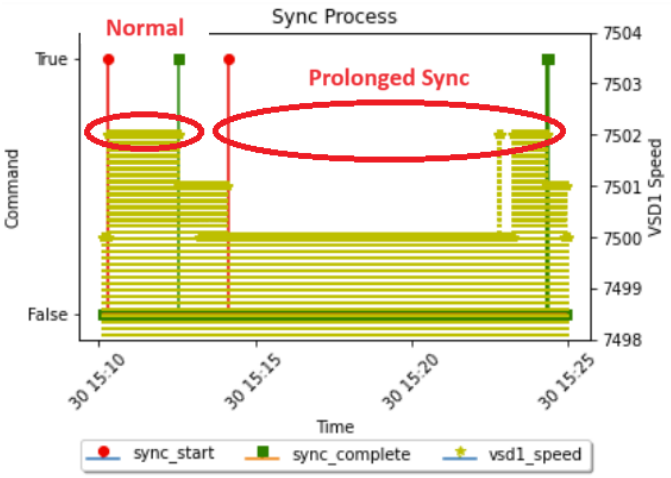

Fig. 9. TDA1: VSD did not accelerate upon start_sync command, causing the synchronization process to take longer than expected.

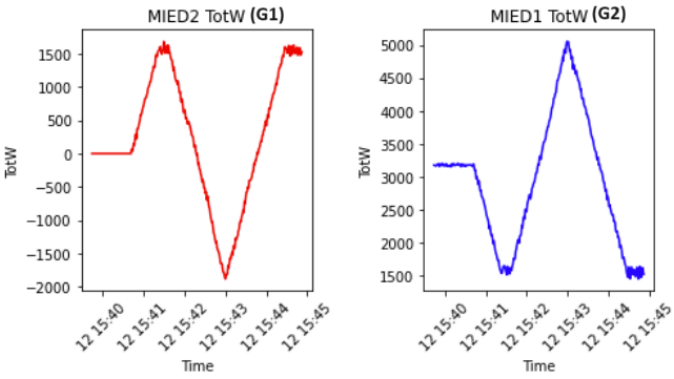

Fig. 10. FDIA3: High reverse power is observed when reverse power protection for G1 is maliciously disabled by attacker and VSD1 speed is set to a value lower than VSD2. The decrease in G1's total real power must be compensated by G2.

ARP spoofing at MSW2 to redirect MMS packets to the attacker's machine. On receiving gen1_p_negative command from MIED2, we modify its value to True. This manipulation tricks the SPLC into believing that reverse power flows have been detected in G1. We then modify the Modbus command at SSW2 to underspin VSD1 to below 7500 (1500 rpm). This sudden decrease in the VSD1 speed triggers a surge in reverse power in G1, eventually causing it to trip and leading to a blackout. As shown in Figure 11, circuit breaker Q2C tripped,

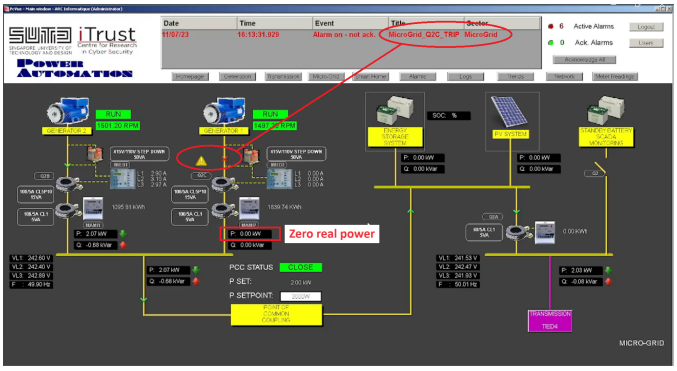

Fig. 11. FDIA4: The attacker underspins the VSD of generator 1, causing it to trip and thereby increasing stress on generator 2.

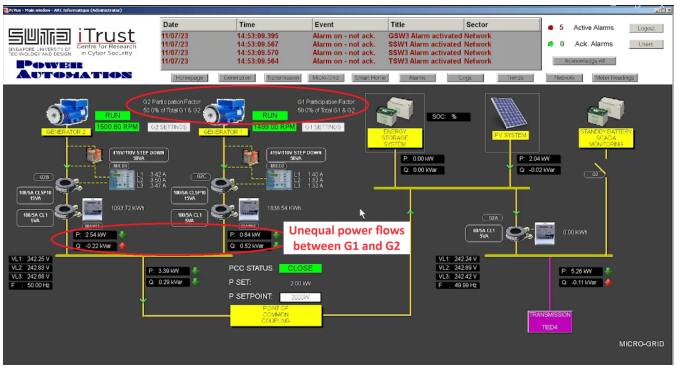

Fig. 12. TDA2: Imbalance power flows from generator 1 and 2 when generator 1 is attacked.

forcing G2 to handle the full demand load.

TDA2: Imbalanced power flows. This attack aims to delay the acceleration of VSD1 when reverse power flows are detected in G1. This attack is achieved by intentionally modifying the gen1_p_negative command from MIED2 to true and matching the speed of VSD1 to the speed of the reference generator (G2). Consequently, there is no speed change to offset the effects of reverse power, causing the transmission lines of G2 to be overloaded while G1 operates below its capacity. Therefore, the grid experiences significant voltage fluctuations, creating imbalances in power flows. The impacts of TDA2 attacks are shown in Figure 12 and Figure 13, where G1 produces less power even though both generators are configured to contribute equally at 50%.

V. Dataset Collection

Each FDIA and TDA experiment is repeated five times under varying demand loads and attack conditions. The demand loads are configured as resistive, capacitive, inductive, or mixed for each experiment. For FDIA2 and FDIA4 attacks, we arbitrarily modify the phase angle and Modbus speed values. For TDA variant attacks, we vary the attack duration and monitor their impacts on grid operations. Each experiment consists of two runs — attack and non-attack — to capture patterns associated with normal and attack behaviors. We use Wireshark to capture the network traffic in PCAP format. Since

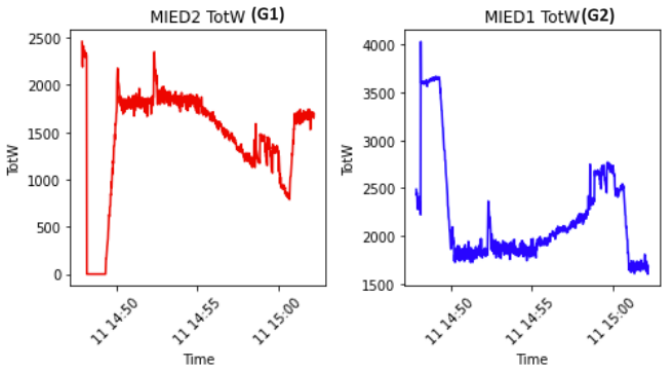

Fig. 13. TDA2: G1's real power exhibits high fluctuations during attack.

TABLE II
THE FINAL FEATURE SET FOR TRAINING MACHINE LEARNING MODELS.

Features	Description
TotW	MMS Measurements from MIED1
TotVar	MMS Measurements from MIED1
Phv.neut.ang	MMS Measurements from MIED1
A.phsA	MMS Measurements from MIED1
sync_start	MMS command from SCADA to initiate synchronization
sync_complete	MMS command from MIED2 indicating complete sync
gen1_p_negative	MMS command from MIED2 to activate reverse power
gen2_p_negative	MMS command from MIED1 to activate reverse power
VSD1 speed	Modbus command from SPLC to regulate VSD1 speed

TotW - Total instantaneous real power
TotVar - Total instantaneous reactive power
Phv.neut.ang - Phase-to-neutral angle
A.phsA - Root mean square (RMS) current phase A

there are four FDIAs and two TDAs, each repeated five times, 30 PCAP files were collected.

From the PCAP files, we analyze the MMS and Modbus packets to identify nine features that best capture the behaviors of generator synchronization and reverse power prevention during attacks. As detailed in Table II, these features consist of four MMS measurements (i.e., TotW, TotVar, A.phsA, and Phv.neut.ang) and five commands (i.e., sync_start, sync_complete, gen1_p_negative, gen2_p_negative, and VSD1 speed). The inclusion of commands in the feature set provides contextual information about the corresponding measurement data. This context is useful for rule-based IDS and machine learning-based IDS. For example, in a rule-based IDS, the sync_start command can be used as a trigger to start monitoring the Modbus speed, while in a machine learning-based IDS, the commands and the corresponding measurements serve as input features for training classifiers to detect anomalies.

A Python program is developed to parse the 9-feature set from PCAPs into CSV format for labeling. The labeling process is based on identifying the onset of an attack condition and labeling the subsequent rows as attacks until the attack ceases. For FDIA1, an attack occurs when two consecutive sync_start = false commands are sent and VSD1 speed = 7500, which indicates a failure to accelerate at 7502. As long as this condition holds, we mark all subsequent rows as attacks. Otherwise, we label the rows as normal.

Following the labeling approach, we generated two labeled CSV datasets, EPIC_A and EPIC_B, each featuring three distinct class labels. Specifically, EPIC_A categorizes data samples into Normal, FDIA, and TDA, while

TABLE III
DETAILS OF THE TWO CSV DATASETS

EPIC_A dataset		
Type	**Label**	**Sample**
Normal	Normal	13448
FDIA1 to FDIA4	FDIA	3356
TDA1 to TDA2	TDA	5555
EPIC_B dataset		
Normal	Normal	13448
FDIA1, FDIA2, TDA1	gen_sync_attacks	5055
FDIA3, FDIA4, TDA2	reverse_power_attacks	3856

EPIC_B classifies samples as Normal, gen_sync_attack, and reverse_power_attack. By differentiating between attack types (i.e., FDIA and TDA) and operation types (i.e., gen_sync_attacks and reverse_power_attacks), we can develop specialized IDS models tailored to each type, enabling a nuanced evaluation of their efficacy. Table III provides more information about the datasets. The corresponding PCAPs and CSV files are available in our GitHub repository [23].

VI. CONCLUSION

The lack of real-world datasets for designing and validating IDS solutions is a significant challenge in smart grid cybersecurity research. This paper addresses this gap by simulating realistic FDIA and TDA attacks on the EPIC testbed, focusing on generator synchronization and reverse power prevention operations. We demonstrated these attacks using ARP spoofing and MITM techniques and open-sourced the datasets on GitHub [23]. These datasets provide a good starting point for the research community to develop and test novel IDS designs for securing power grids, allowing them to replicate attack scenarios and expand our datasets for further research.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

REFERENCES

[1] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC)*, vol. 388, no. 1-29, p. 3, 2016.

[2] V. S. Rajkumar, A. Ştefanov, A. Presekal, P. Palensky, and J. L. R. Torres, "Cyber attacks on power grids: Causes and propagation of cascading failures," *IEEE Access*, 2023.

[3] H. C. Tan, C. Cheh, B. Chen, and D. Mashima, "Tabulating Cybersecurity Solutions for Substations: Towards Pragmatic Design and Planning," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. IEEE, 2019, pp. 1018–1023.

[4] H. C. Tan, C. Cheh, and B. Chen, "Cotoru: automatic generation of network intrusion detection rules from code," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 720–729.

[5] D. Mashima, Y. Chen, M. M. Roomi, S. Lakshminarayana, and D. Chen, "Cybersecurity for modern smart grid against emerging threats," *Foundations and Trends® in Privacy and Security*, vol. 5, no. 4, pp. 189–285, 2023.

[6] M. M. Alani and T. Baker, "A survey of smart grid intrusion detection datasets," in *Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)*. IOS Press, 2023, pp. 5–13.

[7] H. Zeng, Z. W. Ng, P. Zhou, X. Lou, D. K. Yau, and M. Winslett, "Detecting Cyber Attacks in Smart Grids with Massive Unlabeled Sensing Data," in *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2022, pp. 1–7.

[8] J. W. Teo, S. Gunawan, P. P. Biswas, and D. Mashima, "Evaluating synthetic datasets for training machine learning models to detect malicious commands," in *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2022, pp. 315–321.

[9] S. E. Quincozes, C. Albuquerque, D. Passos, and D. Mossé, "Ereno: A framework for generating realistic iec–61850 intrusion detection datasets for smart grids," *IEEE Transactions on Dependable and Secure Computing*, 2023.

[10] M. Conti, D. Donadel, and F. Turrin, "A survey on industrial control system testbeds and datasets for security research," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2248–2294, 2021.

[11] S. Adepu, N. K. Kandasamy, and A. Mathur, "Epic: An electric power testbed for research and training in cyber physical systems security," in *Computer Security: ESORICS 2018 International Workshops, CyberICPS 2018 and SECPRE 2018, Barcelona, Spain, September 6–7, 2018, Revised Selected Papers 2*. Springer, 2019, pp. 37–52.

[12] P. P. Biswas, H. C. Tan, Q. Zhu, Y. Li, D. Mashima, and B. Chen, "A synthesized dataset for cybersecurity study of iec 61850 based substation," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2019, pp. 1–7.

[13] P. P. Biswas, Y. Li, H. C. Tan, D. Mashima, and B. Chen, "An attack-trace generating toolchain for cybersecurity study of iec61850 based substations," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–7.

[14] P. Radoglou-Grammatikis, V. Kelli, T. Lagkas, V. Argyriou, and P. Sarigiannidis, "Dnp3 intrusion detection dataset," https://dx.doi.org/10.21227/s7h0-b081, 2022.

[15] W. Lin, M. R. Saifuddin, and B. Chen, "The design and implementation of a cyber exercise on epic microgrid testbed," in *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2023, pp. 1–7.

[16] S. iTrust, "About critical infrastructure defence exercise 2022," https://itrust.sutd.edu.sg/cidex-2022/, 2022, 2024-03-04.

[17] C. M. Ahmed and N. K. Kandasamy, "A comprehensive dataset from a smart grid testbed for machine learning based cps security research," in *Cyber-Physical Security for Critical Infrastructures Protection: First International Workshop, CPS4CIP 2020, Guildford, UK, September 18, 2020, Revised Selected Papers 1*. Springer, 2021, pp. 123–135.

[18] A. Mumrez, M. M. Roomi, H. C. Tan, D. Mashima, G. Elbez, and V. Hagenmeyer, "Comparative study on smart grid security testbeds using mitre att&ck matrix," in *2023 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2023, pp. 1–7.

[19] M. R. Saifuddin, L. Wei, H. C. Tan, and B. Chen, "Coordinated Network Attacks on Microgrid Dispatch Function: An EPIC Case Study," in *European Symposium on Research in Computer Security*. Springer, 2022, pp. 26–45.

[20] M. M. Roomi, P. P. Biswas, D. Mashima, Y. Fan, and E.-C. Chang, "False Data Injection Cyber Range of Modernized Substation System," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 2020, pp. 1–7.

[21] M. M. Roomi, S. S. Hussain, D. Mashima, E.-C. Chang, and T. S. Ustun, "Analysis of False Data Injection Attacks against Automated Control for Parallel Generators in IEC 61850-based Smart Grid Systems," *IEEE Systems Journal*, 2023.

[22] "IEC 61850 - Communication networks and systems for power utility automation - ALL PARTS," Feb 2020. [Online]. Available: https://webstore.iec.ch/publication/6028

[23] IARCS, "EPIC_Attack_Datasets," April 2024. [Online]. Available: https://github.com/smartgridadsc/EPIC_Attack_Datasets