Cyber-physical metropolitan area digital substations test bench for evaluating intrusion detection systems

Santiago Sanchez-Acevedo\textsuperscript{1}, Tesfaye Amare Zerihun\textsuperscript{1}, Hristo Koshutanski\textsuperscript{1}, and Alejandro Garcia-Bedoya\textsuperscript{2}

\textsuperscript{1}Department of energy systems, SINTEF Energy Research Trondheim
\textsuperscript{2}Cybersecurity unit Eviden BDS R&D Madrid

June 07, 2024
Abstract—This paper introduces a cyber-physical test bench for metropolitan area digital substations, which serves as a platform for studying advanced cyber-attacks and evaluating intrusion detection systems. The test bench provides a hardware-in-the-loop environment with a high Technology Readiness Level (TRL). Additionally, the paper demonstrates the testbed’s capabilities by testing and validating an AI-based intrusion detection system developed by an industrial company. The demonstration includes testing the cybersecurity tool against a wide range of cyber-attacks, such as false data injection, packet replay, and time desynchronization, on relevant substation automation protocols commonly used in EU power grid infrastructures like IEC-60870-5-104 and IEC61850. The results of the demonstration indicate that the test bench effectively simulates realistic impacts on substation operation when subjected to attacks, thus providing the opportunity to validate the anomaly detection solution.

I. INTRODUCTION

Electric Power Systems (EPS) play a key role on economy development at regional and national levels and are considered critical infrastructure. Nowadays, EPS are increasingly using information and communication technology (ICT). This integration of ICT in EPS can be defined as cyber-physical power systems (CPPS). A CPPS consists essentially three layers: physical layer, communications layer and control/decision layer. The physical layer comprises generation systems, transmission/distribution lines, substations and consumer loads. Meanwhile, the communications layer may consist of either local area networks within substations or wide area communication systems utilized by system operators. The decision layer of a Cyber-Physical Power System (CPPS) involves the control center of system operators which serves as the central hub for managing and controlling the EPS [1]–[3]. With the addition of the cyber-communication layer to the EPS, the vulnerability increases. A potential cyber attack on the ICT could result a blackout which has a potential to jeopardize regional or national economic development. Therefore, it is imperative to study cascading effects of attacks to this critical infrastructure. Researchers have been developing cyber-physical testbeds of digital substations of EPS to study cyber attacks [4] using software define networking characteristics to block the attack. Some authors have presented synthetic simulated traffic of digital substations in [5]. Some examples of cyber-attacks on CPPS are listed in [6].

Very few intrusion detection datasets have focused on attacks on IEC 61850 protocol, such as for GOOSE [7] and SV [8]. There is a pressing need to extend and complement those with additional attack datasets to allow for enriched anomaly and intrusion detection. In addition, high TRL (TRL7) testbench environment offers more advanced feature engineering analysis on the realistic traffic footprint of these protocols, and subsequently more accurate validation of IDS solutions. The work in this paper complement the OT protocol intrusion detection datasets [7], [9] with additional attacks of high impact and more hidden traffic into existing communication sessions to detect which also represent a threat to digital substation operations.

Main contribution of this paper is the testbench modelling the cyber-physical metropolitan area digital substations which can be used for studying advanced cyber-attacks and evaluating intrusion detection systems. By providing a hardware-in-the-loop environment, the testbench offers a realistic setting with a high technology readiness level (TRL), incorporating essential digital substation devices from reputable vendors.

II. DISTRIBUTION SYSTEMS SCADA AND DIGITAL SUBSTATIONS

a) Digital substation equipment: Modern substations follow IEC 61850 standard. IEC 61850 facilitates communication of the devices used on tasks like measuring, monitoring, protection and control in the digital substation (DiSt). Therefore, key devices used in a digital substation are the merging unit (MU), a device that measures voltage and currents and streams them in the form of digital sampled values in the local area communication network of the DiSt. An intelligent electronic
device (IED) is an advanced device with monitoring and protection features and it is able to communicate to a SCADA system. Additionally, the DiSt switches make possible traffic at process bus between MU and IEDs.

b) Digital substation to control centre: Metropolitan area communications from the digital substations to the distribution system operator’s control centre can be achieved by implementation of IE 60870-5-104 (IEC-104).

III. INTRUSION DETECTION FOR DIGITAL SUBSTATIONS

In the last few years, a considerable progress has been made on intrusion detection in ICS [10] and critical infrastructure protection [11]. Intrusion detection systems (IDS) are largely based on signature or anomaly detection. The strength of signature-based IDS is on the use of knowledge of attack signatures to offer high precision and accuracy for the detection of the signatures. This type of IDS has been used for many years now and have showed that under specific requirements of timely signatures identification and updates, operators of power substations can achieve a good protection of their assets. The signature-based IDS offer low rate of false positives and high precision, but are subject of false negatives, that is the difficulty or uncertainty of detection of variations of attack signatures or new attacks in a given domain, also referred to as zero-day attacks. On the other hand, the field of anomaly detection has also evolved for several years now, especially in the domain of cybersecurity for power grid [12]. The underlying principle of anomaly detection is on the modelling and learning of baseline behavioural patterns of an environment normal operations, and detect any anomalies as significant deviation from the baseline patterns. Anomaly detection is subject of higher rate of false positives especially in an environment of frequent or dynamic changes of baseline behavioural patterns, but offers a solid detection of attacks either known or unknown. In this work, we deliberately decided to go in the direction of anomaly-based IDS as an adequate means for cybersecurity monitoring of power grid digital substations. The main arguments for our decision are:

1) Industrial automation protocols such as IEC-104, DNP3, Modbus, and IEC61850 widely used in PowerGrid substations exhibit specific network behavioural patterns with consistent behaviour over time, which facilitates learning and minimises the potential false positive rate that AD IDS are often subject of.

2) Anomaly detection allows to model the baseline behaviour of the ICS/SCADA protocols with specific protocol-level behavior features. Protocol-specific behavioral features are essential to ensure visibility and insight into protocol communications, to model and learn behavioral patterns and achieve high performance IDS. Refer for instance to the datasets and their features [9], [13], [14].

3) Unsupervised Deep Learning algorithms, such as Autoencoders (AE), have shown a solid basis for network anomaly detection, as well as for the detection of advanced and highly persistent cyber-attacks with no prior knowledge of such [15].

We will overview related work on IDS solutions specific to digital substations industrial protocols. Radoglou et al. [16] presented an anomaly detection system for IEC-104 based SCADA systems using One-Class Support Vector Machine (OC-SVM). The OC-SVM results in a binary function, which captures regions in the input space where the probability density of the data lives. The subsequent work [17] proposed an IDS for the DNP3 protocol called DIDEROT, which consists of two layers of detection: the first is a decision-tree classifier to identify certain attacks with DNP3 protocol and a second layer of detection using an AE algorithm capable of detecting anomalies. This work is aligned with our decision on anomaly detection using AE but we postulate to perform the anomaly detection before the classification in order to limit potential false negatives or false positives and respect the specificity of baseline behaviour of a digital substation environment.

A work by Alabe et al. [18] proposed a two-stage deep learning approach using an AE and long short-term memory (LSTM) to detect anomalies in Electric Power Steering (EPS) system’s sensor data. This work inspires to adopt LSTM as a more advanced form of NN with respect to the Recurrent Neural Network (RNN). However, we observed that LSTM increased on average 5x times complexity in detecting anomalies but added little value (with no noticeable increase of accuracy) to the detection of anomalies with respect to the use of RNN. The work by Kravchik et al. [19] demonstrated the detection capabilities of algorithms such as Convolutional Neural Networks and AE on industrial protocols. In addition, the use of PCAs is proposed to reduce the possibility of Adversarial Attacks. Our approach also uses the AE in different industrial protocols, but we do not use PCAs to reduce the dimensionality since the AE itself reduces the dimensionality to the prefixed dimension of the latent space.

Song et al. [15] presented a study on the results obtained by applying AE for Network Intrusion Detection in zero-day attacks. It shows the potential of AE for detection of intrusions especially for the targeted by us zero-day attacks. It however identifies the importance (and the difficulty) in identifying the optimal architecture of AE. It shows how the architecture of the AE model and the size of the latent space affect the performance of the network IDS, and concludes that the number of hidden layers does not impact that much on the accuracy of anomaly detection but the dimension of the latent space is the key factor.

In our experiments, we used a dynamic auto configurable AE architecture to maximise the learning of network behaviour, and to reduce the reconstruction loss. For the different OT protocols, we devised different dimensionalities of behavioural features specific per each protocol, and a specific DL models was produced to maximise efficiency in detection. However, despite the advancement of our anomaly-based intrusion detection system, we identified the need of a realistic testbed environment for validation and demonstration of the
cyber-attack execution, impact and detection. The proposed wide area digital substation testbench has shown a suitable basis to perform such validation activities. This is also motivated by the fact that despite the recent release of some intrusion detection datasets such as for IEC-104 [9], DNP3 [13] and Modbus [14], although interesting for the attacks they capture, they do not sufficiently represent a realistic digital substation environment with real operational network traffic to validate anomaly-based IDS for high impact and persistent cyber-attacks.

IV. CYBER-PHYSICAL POWER SYSTEM

This section describes the testbench developed to recreate a CPPS with the most relevant devices available in a real distribution power system environment.

a) Distribution network: An extended version of medium voltage distribution network CIGRE MV benchmark has been used to represent the metropolitan area distribution power system. Figure 1 shows the network topology used in this paper.

b) Communications and monitoring system: The setup shown in Figure 1 is used to model the extended CIGRE MV benchmark system in the National Smart Grid Laboratory in Trondheim, Norway. This setup employs a hardware-in-the-loop configuration, integrating both physical hardware and emulated measurement and control devices together with real-time simulator for modelling the distribution grid.

Three different configurations of digital substations are considered. The first configuration (Substations 1 and 2) assumes the utilisation of Remote Terminal Units (RTUs) to directly read breaker status from virtual MU and ED using IEC61850 protocol and exchange data with control room with IEC-104 protocol. The second configuration (Substation 3) has IED to exchange IEC-104 traffic to control centre. The third configuration (substation 4) consists one RTU and IEDs, where IEDs are assumed to work in local operations such as protection on the process bus using IEC61850 and RTU is used for communicating towards the remote control centre using IEC-104.

The first two substations (1 and 2) are emulated using OPAL RT real time simulator and RTUs SIMATIC 3030C, while Substation 3 and substation 4 are modelled using hardware RTUs and IEDs from different vendors. Substation 3 is modelled with SIPROTEC 7SJ85 IED from SIEMENS where as substation 4 is modelled with two devices; one from ABB RET670 and another from SIEMENS – 6MU85. Substation 4 also has a remote terminal unit (RTU) from SEIMENS – one SICAM A8000. The power system distribution network is emulated using OPAL-RT real time simulator.

The communications at the substation level is based on the IEC 61850 standard. Sampled value traffic within each substation is generated using virtual merging units emulated inside the OPAL-RT simulator. IEDs in substation 3 and 4, subscribe to the sampled values (SVs) of their respective process busses. Subsequently, RTUs and IEDs exchange data with the operator’s SCADA software using the IEC-104 protocol.

The switches model used for the process bus are MOXA-pt-7728 series, and for the station bus is Planet IGS-6325-16P4S. These are standard widely used industrial switches.

AVEVA CITEC SCADA software is used to emulate the central control room. All the virtual IEDs, MUs and hardware IEDs were synchronised using precision time protocol (PTP) with a satellite synchronised network clock SEL-2488. Furthermore, industry grade switches such as MOXA 7728 series are used for local communication inside a substation.

V. EXPERIMENTAL RESULTS

This section presents results obtained from the use of cyber-physical testbench for validating a cyber-securety intrusion detection method.

A. Cyber-security Scenarios

In contrast to the known intrusion detection datasets, refer to Section III, we focused on complementary attack scenarios that are more difficult to detect, leave minimum traffic footprint, and, at the same time, generate high impact on the substation operation. We define the following adversary model for all attack scenarios: The adversary has a full knowledge and expertise of ICS/SCADA protocols, is able to sniff (observe) traffic on the switch level, and is able to inject traffic (spoofed packets) at process, station buses and wide area communications. We exclude from the model that the adversary has control over the network infrastructure such as industrial switches used for station and process bus communications. It is the underlying assumption that the IDS receives a correct mirroring of the systems traffic.

False Data Injection. After observing substation traffic, an attacker’s script synchronises with a valid communication session and injects false packets with crafted payload values. The impact of this attack scenario is high as it aims to introduce false measured values or unauthorised function code (command) execution as communicated within a valid-looking protocol session. Traditional Packet-based
IDS systems such as Snort\textsuperscript{1} or Suricata\textsuperscript{2} will need special configuration and knowledge (if possible at all) to detect such attacks. For instance, in the case of IEC-104, we introduced an operator of the testbed to perform some routine operations due to maintenance of opening and closing a breaker, where the attacker (through its malware script) observes the commands used to open and close the breaker (C\_DC\_NA\_1), and then uses well crafted packets to inject into the valid communications such commands to manipulate breaker’s state. In this case, although the spoofed packets represent a valid-looking IEC-104 packet, they cause an anomaly on a flow level and on the behaviour from the SCADA HMI to the RTU. Similar attack scenarios were implemented for the three protocols IEC-104, IEC61850 GOOSE, and SV with varying packet injection frequency.

**Packet replay.** After observing substation traffic, the attacker’s script replays selected packets from past sessions between the same devices to introduce discrepancy on the logic of the RTUs/IEDs. The difference from the FDI attack is that there is no session injection, it is easier to distinguish spoofed packets by RTUs/IEDs, and attack impact has a more limited effect on the RTUs/IEDs. This attack scenario was implemented for the protocols IEC61850 GOOSE and SV with varying number of replacted packets. These protocols are layer 2 of the OSI and exhibit a simpler session information.

**Time desynchronisation.** The attacker observes the process bus communications with the aim to introduce and inject packets from a bogus entity announcing its role as a master clock to RTUs/IEDs to make them accept this role and attempt to sync their internal clock with the new bogus entity. The time desynchronisation attack has a direct high impact on GOOSE and SV communications as it leads to rejection of packets, kind of a denial of service (DoS) attack due to the loss of precision. In many industrial devices setups, remote time synchronisation is main decision of accepting measured values. The attack scenario was implemented for the Precision Time Protocol (PTP) widely used in digital substations.

**B. Intrusion detection based on AI-behavioural analysis**

The tool used for validation of the attacks on the test-bench is an AI-based anomaly detection IDS developed by Eviden BDS R&D Spain, and matured under the ELECTRON\textsuperscript{3} project in the power grid resilience domain.

This tool is specialised in detecting behavioural-based anomalies and intrusions in ICS/SCADA protocols. It learns the normal pattern of a given SCADA communication environment to establish a baseline of legitimate behaviour. The tool detects any anomaly as deviation from this baseline. It uses a rich set of features of network traffic of the OT protocols mentioned above that facilitates the detection of cyber-attacks by the machine leaning model. The tool is based on Autoencoder to model patterns of benign traffic and identify anomalous behaviour.

Figure 2 shows the high level view of the tool’s workflow. The initial phase involves training deep learning models using legitimate traffic. This process can be conducted either online or offline. Once the model is trained, the tool switches to monitoring or inference mode. The first module is OT-FlowMeter (OTFM), which feeds the Brain module with a set of network behavioural features extracted from the monitored network traffic. The OTFM is based on the well known open-source community CICFlowMeter\textsuperscript{4} tool but substantially customised to extract more network traffic features both generic and OT protocol specific necessary to detect anomalies.

There is a Brain module that contains deep learning models one per each OT protocol family. It processes traffic collected by the OTFM for training the Autoencoder algorithm. The Brain detects possible deviation from the pattern learned during the training phase. Through the utilization of error reconstruction, coupled with a predetermined threshold, it classifies incoming traffic into legitimate or anomalous, and provides additional explainability information to understand the nature of the anomaly. For instance, what features have been the most critical for detection, and gives evidence showing the deviation of the anomaly from the normal training data.

**C. Cyber attacks considered**

1) Attacks on IEC-104:

a) FDI Monitoring Direction: Injection of spoofed IEC-104 packets of single point information ASDU type 0x01 M\_SP\_NA\_1 within a valid TCP session from RTU to SCADA. Attacker’s device can monitor interrogation commands from SCADA to RTUs and inject few milliseconds earlier a well spoofed IEC-104 packet (matching Rx, Tx session counters) before the original one. IMPACT: The attacker achieves a persistence in falsifying data so that SCADA believes that these are values being received by the RTU.

b) FDI Control Direction: Injection of spoofed IEC-104 packets of type C\_DC\_NA\_1 within a valid TCP session from SCADA to RTU. Attacker’s device can monitor communications from SCADA to RTUs and injects few milliseconds earlier a well spoofed IEC-104 packet (matching Rx, Tx) before the original one. IMPACT: The attacker succeeds to execute the double command C\_DC\_NA\_1 for manipulating the breaker state at the RTU and manipulates the process.

\textsuperscript{1}https://www.snort.org
\textsuperscript{2}https://suricata.io
\textsuperscript{3}https://cordis.europa.eu/project/id/101021936; https://electron-project.eu
\textsuperscript{4}https://github.com/ahlashkari/CICFlowMeter
2) Attacks on IEC-61850 GOOSE:

a) Packet replay: Replay GOOSE packets of previous valid state communications (e.g., status or control of a breaker) simultaneously to a current state communication. The attacker attempts to manipulate breaker state reusing previous valid packets. The attack rate varies from 4 packets to 100 packets per session injection.

b) FDI version 1: Injection/broadcasting of GOOSE packets like a real GOOSE engine where new stNum is announced, reset sqNum and new time stamp (T) for the new state simulated. The new stNum is incremented following the original stNum, and sequentially increments the sqNum and updates its T. In fact, this attack aims to achieve a persistence offering a real GOOSE message sequence for all the time, each attack has been streamed. IMPACT: Manipulated system state and operation: 1) with spoofed stNum and sqNum, and changes the values inside GOOSE’s data unit. 2) Attacker enters into a GOOSE heartbeat rate, synchronises with that rate, and injects 3 packets in the rhythm of heartbeat, and becomes silent for a period of time when it repeats the injection of new packets.

c) FDI version 2: The attacker observes the GOOSE heartbeat, enters into the rhythm of packets announcements, and starts injecting false packets few milliseconds before the original ones following the same frequency. The attacker stays in the network for as long as needed without disconnecting, thus sustaining false packets injections towards the IEDs.

3) Attacks on IEC-61850 SV:

a) Packet replay: Capture and replay SV frames with values from previous communications. The attack rate was 1 frame replayed each 500ms.

b) FDI: Injection of SV frames with false values and spoofed smpCnt values: The attacker synchronises with the current multi-casting state of the SV frames, and then awaits when the smpCnt resets upon the 1 s clock duration, then the attacker begins to inject false SV frames with a new counter from 0 to 3999 for 1 s duration, following what the SV standard mandates. IMPACT: Manipulate system state and operation: Figure 3 shows how the IED becomes unstable when receiving multiple frames one set with spoofed values by the attacker and the original set with different values. The device resets the estimated values to 0.0 kV as an effect of this attack. This attack can persist for a long time. Once it is over, the device resets the values to those from the original frames.

4) Attacks on PTP (IEEE 1585):

a) Desynchronization of RTUs/IEDs: The target was to disrupt subsequent GOOSE and SV-based communications due to a low precision clock source. An attacker introduces a bogus entity announcing its role as a master clock role to the original master clock make him accept this role. Therefore, RTUs/IEDs attempt to sync its internal clock with the new rouge entity. IMPACT: the IEDs were not correctly synchronised for a period and then stopped remote sync and switched to a local sync with much lower precision. Other anomalies were detected as a result of this attack such as the main (original) master clock first got shorter duration of communications and announcements, and then moved to inactivity for a period of time. This was another sign of detection the attack by detecting idle or inactivity state of the master clock.

D. Test-bench validation of intrusion detection

We assessed the effectiveness of the anomaly-based IDS tool using metrics of Accuracy, F1-Score and confusion matrix.

a) Confusion Matrix: It is common to use a confusion matrix to understand the behaviour of a deep learning algorithm. The confusion matrix is characterised by comparing the actual values with the values predicted by the algorithm. In a binary classification problem, the confusion matrix is described with (1).

\[
\begin{array}{c|c|c|c}
\text{Actual Neg.} & \text{Predicted Negative} & \text{Predicted Positive} \\
\hline
\text{TN} & \text{FP} & \text{TP} \\
\hline
\text{FN} & \text{FN} & \text{FP} \\
\end{array}
\]

where, TN (True Negatives) are the number of legit flows that have correctly been predicted as legit traffic. FP (False Positives) are the number of legit flows that have wrongly been predicted as attack traffic. FN (False Negatives) are the number of attack flows that have wrongly been predicted as legit traffic. TP (True Positives) are the number of attack flows that have correctly been predicted as attack traffic.

b) Accuracy: known as the ratio between the correct predictions divided by the total number of predictions (2).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

where, TP and TN are the number of attack and legit flows correctly predicted. FN and FP are the number of attack and legit flows wrongly predicted.

c) F1-Score: indicates the harmonic mean of precision and recall, giving equal weight to both measures. It is particularly useful when there is an uneven class distribution or when false positives and false negatives have different costs.

\[
F1 - \text{Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \quad \text{and} \quad \text{Recall} = \frac{TP}{TP + FN}
\]

The results obtained for different attacks are shown in Figure 4 and Figure 5. As can be seen, the results obtained are promising for the IDS tool. In all attack cases, an accuracy over 92% was achieved, with an F1-Score for legitimate
cases exceeding 94% and an F1-Score for anomalous cases surpassing 85%. Furthermore, the confusion matrices confirm the good results obtained. In some cases, there are errors in predicting legitimate traffic as anomalous traffic, but this is an acceptable error. Given that the opposite error, misclassified anomalous traffic as normal traffic, is the most critical error for anomaly identification systems.

The high detection rate of metrics shown in Figures 4 and 5 is explained by the use of specific features defined for the OT protocols that allow to observe desired deviations on core protocol fields, such as volumetric and statistical features of function code utilisation, control code control utilisation, or other protocol-specific session features such as state number, protocol timestamp, sequence number, sample counter, etc. For instance, in the case of IEC-104, there are more than 1800+ features extracted from the protocol communication, and 400+ for the case of GOOSE and SV.

VI. CONCLUSIONS AND FUTURE WORK

The developed test-bench is a realistic digital substation environment corresponding to TRL7. This environment implements relevant substation automation protocols widely used in EU power grid infrastructures such as the IEC-104 and IEC61850. Selected attacks of high impact on these protocols have been executed and captured to support the community of researchers and practitioners in validation of their IDS solutions. The attacks shown a realistic impact on the substation operation. Another added values of the TRL7 test-bench is the possibility to validate anomaly detection solutions on the moulding capacity of the baseline traffic footprint of this environment. For instance, IEC61850 SV requires capacity of 4000 pps for each one second elapsed time, with several devices being connected. Future work will focus on extending the testbench to cover other SCADA protocols such as IEC61850 MMS and C37.118 to support validation of IDS solutions on targeted cyber-attacks on these protocols.

REFERENCES


