Semi-supervised deep learning-driven anomaly detection schemes for cyber-attack detection in smart grids

Abdelkader Dairi, Fouzi Harrou, Benamar Bouyeddou,
Sidi-Mohammed Senouci, and Ying Sun

[1] Abdelkader Dairi Laboratoire des Technologies de l'Environnement LTE, BP 1523 Al M'naouar ENP Oran, University of Science and Technology of Oran-Mohamed Boudiaf (USTO-MB), El Mnaouar, BP 1505, Bir El Djir 31000, Algeria
abdelkader.dairi@univ-usto.dz

[2] Fouzi Harrou, Ying Sun CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
fouzi.harrou@kaust.edu.sa, ying.sun@kaust.edu.sa

[3] Benamar Bouyeddou STIC Lab, Department of Telecommunications, Abou Bekr Belkaid University, Tlemcen, Algeria
and LESM Lab., University of Saida-Dr Moulay Tahar, Department of Electronics, Faculty of Technology, Saida, Algeria benamar.bouyeddou@univ-saida.dz

[4] Sidi-Mohammed Senouci DRIVE Laboratory, University of Burgundy, Nevers, France, Sidi-Mohammed.Senouci@u-bourgogne.fr

Abstract. Modern power systems are continuously exposed to malicious cyber-attacks. Analyzing industrial control system (ICS) traffic data plays a central role in detecting and defending against cyber-attacks. Detection approaches based on system modeling require effectively modeling the complex behavior of the critical infrastructures, which remains a challenge, especially for large-scale systems. Alternatively, data-driven approaches which rely on data collected from the inspected system have become appealing due to the availability of big data that supports machine learning methods to achieve outstanding performance. This chapter presents an enhanced cyber-attack detection strategy using unlabeled data for ICS traffic monitoring and detecting suspicious data transmissions. Importantly, we designed two semi-supervised hybrid deep learning-based anomaly detection methods for intrusion detection in ICS traffic of smart grid. The first approach is a Gated recurrent unit (GRU)-based stacked autoencoder (AE-GRU), and the second is constructed using a generative adversarial network (GAN) model with a recurrent neural network (RNN) for both generator and discriminator that we called GAN-RNN. The employment of GRU and RNN in AE and GAN models is expected to improve the ability of these models to learn the temporal dependencies of multivariate data. These models are used for feature extraction and anomaly detection methods (Isolation forest, Local outlier factor, One-Class SVM, and Elliptical Envelope) for cyber-attack in power systems. These approaches only employ normal events data for training without labeled attack types, making them more attractive for detecting cyber-attack in practice. The detection performance of these approaches is demonstrated on IEC 60870-5-104 (aka IEC 104) control communication that is often utilized for substation control in smart grids.

Keywords: Cyber-attack detection, Protocol IEC 104, deep learning, semi-supervised methods, anomaly detection.

1 Introduction

Smart grids play a central role in the efficient management and control of the produced energy. They are designed to transform traditional electricity grids from totally centralized and isolated systems to fully connected and decentralized systems that rely on distributed generation, transmission, distribution, and monitoring processes [1]. Substantially, modern power grids come up with many enhancements, including self-healing, offering more services to consumers (e.g., consumers with a multitude of services), improving power quality and reliability, integrating different sources of energy, and enabling robust and automatic control and supervision procedures [2, 3]. Power grids have become more dependent on common information technology computing and networking infrastructure for conducting all aspects of operation and maintenance, which significantly increased their vulnerability to anomalies and cyber-attacks. Since intra-SCADA communications are based principally on DNP3 (Distributed network protocol DNP3), International Electro-technical commission IEC60870-5, Profibus IEC 61850 [1], the backbone network involves different wired and wireless technologies and protocols including Zigbee, WiFi, IP-based networks and cellular networks (e.g., 4G) [4]. However, with the evolving of power grids into a cyber-physical systems, their vulnerabilities to cyber-attacks increased more than before [5]. Generally speaking, cyber-attacks in power systems could result in the loss of availability and may have a real severe impact on physical processes (e.g., human life, environment, damage of equipment, etc.), loss of productivity, and revenue.

Power grids are generally placed in remote sites, and they are remotely monitored and controlled via SCADA (Supervisory Control and Data Acquisition) Systems [6, 7]. The central role of SCADA systems is gathering and analyzing data, communicating, and controlling the operation of systems in real-time [8, 9]. Essentially, SCADA systems comprise several key components, including the Human Machine Interface (HMI), the Master Terminal Unit (MTU), and Remote Terminal Units (RTU) [9, 10]. Specifically, HMI enables operators to monitor the inspected process's state and adjust its control settings. The role of the MTUs, the heart of the SCADA system, is to store and process information gathered from field devices and communicate control signals. RTU receives commands from the MTU to control the local process, acquire data from field devices, and continuously transmit it to the MTU. Each RTU is usually connected to numerous sensors and actuators managing a local process or field equipments. The communication between these components can be wired or wireless via the internet. The common protocols (ModBus, Profibus, DNP3) used in the communication between these components present many vulnerabilities regarding the authentication mechanisms between the MTU and the other components, the integrity of the transmitted packets, and the anti-replay mechanisms. Modern power systems are continuously exposed to malicious cyber-attacks and anomalies, which are challenging and hard to identify [11–13]. If not detected accurately and timely, cyber-attacks and anomalies in industrial systems could decrease plants' productivity, efficiency, and safety, cause severe economic losses, and damage the attacked system. For example, disruption of electric power operations could significantly impact national security and the economy. Therefore, designing an accurate and sensitive intrusion and anomaly detection method is undoubtedly necessary to ensure the productivity and safety of power systems against cyber-attacks and anomalies.

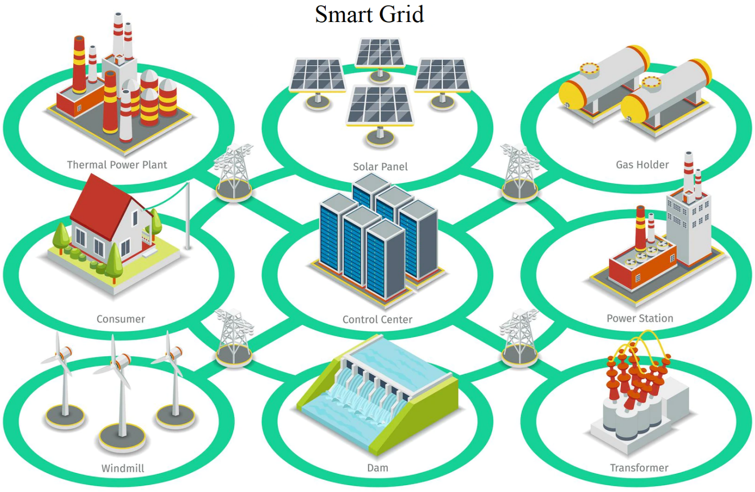

Fig. 1: Smart grid

Numerous intentional cyberattacks targetting power systems have occurred in the last two decades [14, 15, 5]. For example, in December 2015, a cyber-attack on the Ukraine power grid caused circuit breakers at 30 substations to trip, cutting power to around 225,000 customers [16]. A denial of service (DoS) targeted the telephone system and communication network, making the call center unavailable to customers. In addition, the malware implanted on the human-machine interface (HMI) was employed for deleting software on the system, which prevented the operator from characterizing the extent of the power outage and hampered repair actions. SCADA equipment was rendered inoperable, and power restoration had to be completed manually. It could have resulted in severe damage to the power grid. These attacks highlight the lack of a security-driven technique for building up and maintaining power systems. Hackers exploit vulnerabilities in these critical systems to penetrate and gain access to SCADA networks that monitor physical processes, collect, manipulate and destroy critical data exchanged between facilities and operators, and implant malicious malware to disrupt the normal operating conditions of systems [17]. Moreover, attacks can be implemented on different layers (networks, mac, and physical) by exploiting the hosted protocols (e.g., DNP3, Modbus, Transmission control protocol TCP, User datagram protocol UDP, Internet control message protocol ICMP, and Hypertext transfer protocol HTTP) [1].

Power grids are exposed to a variety of cyber-attacks, which can be grouped into three groups according to the attacker's targets [18, 2]. Attacks within the first group attempt to violate confidentiality and potentially privacy by infiltrating the system and gaining access to devices and stored/exchanged data [18]. Hence, they can collect, manipulate, change and exploit information about the grid's operation and consumers (personal information, consumption profile, bills). To this end, various techniques are elaborated, including traffic sniffing, eavesdropping, man-in-the-middle-based attacks, and IP (Internet protocol) / ARP (Address resolution protocol) spoofing [2]. The second group of attacks consists of data integrity attacks (DIA) [19]. In DIA attacks, malicious attackers try to falsify the sensor measurements by compromising device settings and commands and injecting false data [20, 21]. The third group relies on the availability, and their main objective is to make targeted devices and data inaccessible temporarily or permanently and at least delay their responses. They can manifest in different forms, like link failures, bandwidth exhaustion, flooding, and malformed data structure [22]. They are commonly referred to as DOS (Denial of service) attacks and DDOS (distributed DOS) if initiated using multiple sources. In practice, several techniques have been used to create DOS situations, including Jamming, buffer overflow, TCP-based floorings, UDP-based amplification, and malwares [23, 24].

Smart grid networks play a central role by ensuring the distribution and transmission of electric supply to consumers. Connecting smart grids with the internet generated much space for various types of anomaly injection and cyber-physical attacks. In addition, numerous industrial control protocols, such as Modbus, Goose, or IEC 104, are now used to operate within a smart grid substation and interconnect the control system with power equipment. However, this makes a smart grid network a potential target for outside hackers. Protecting smart grid networks is indispensable to ensuring security and energy production, and it has recently become even more critical than ever before [16]. This is a challenging task because smart grid networks are continuously exposed to malicious attacks coming from the outside and inside the network.

1.1 Contribution

This chapter presents a semi-supervised deep learning-based approach for detecting suspicious communications in ICS/SCADA networks. Unlike supervised methods, semi-supervised cyber-attack detection methods need only the data of normal events to train the detection model, making them more attractive for detecting cyber-attacks in smart grid networks since it is not always easy to get accurately labeled data. Of course, the major contributions of this chapter are summarized bellow.

In this chapter, two effective deep learning-driven cyber-attack detection schemes have been introduced for ICS traffic monitoring and detecting suspicious data transmissions in a smart grid network. Importantly, the proposed approaches combine the advantages of deep learning models and semi-supervised anomaly detection methods for intrusion detection in ICS traffic of smart grid. It is worth noting that these semi-supervised anomaly detection methods need only normal instances of ICS traffic (i.e., unlabeled data), which make them appropriate for detecting unknown attacks. We used the proposed deep learning to learn important features in ICS communication traffic data and the sensitivity of semi-supervised anomaly detection methods for cyber-attack detection. The first approach is a Gated recurrent unit (GRU)-based stacked autoencoder (AE-GRU), and the second is constructed using a generative adversarial network (GAN) model with a recurrent neural network (RNN) for both generator and discriminator that we called GAN-RNN. The employment of GRU and RNN in AE and GAN models is expected to improve the ability of these models to learn the temporal dependencies of multivariate data. These models are used for features extraction and anomaly detection methods (Isolation forest, Local outlier factor, One-Class SVM, and Elliptical Envelope) for cyber-attack in power systems. As we know, this is the first work applying semi-supervised deep learning-driven anomaly detection algorithms to detect attacks in ICS flow data of smart grids. At first, the considered deep learning-based models are built based on training data (normal ICS traffic) and then used to detect cyber-attacks that can be launched from inside or outside the network. We assessed the effectiveness of these approaches on IEC 60870-5-104 (aka IEC 104) control communication that is often utilized for substation control in smart grids. Four statistical indices are employed to compare the discrimination accuracy of the considered methods: accuracy, precision, F1-score, and the Area Under the Curve (AUC). Results revealed the promising performance of the proposed approaches in detecting different types of attacks.

Section 2 highlights literature reviews on the related works, and Section 3 introduces the proposed deep learning-based malicious attack detection methods. Section 4 assesses the proposed method on IEC 60870-5-104 (aka IEC 104) control communication that is commonly utilized for the substation control in smart grids. Finally, Section 5 concludes this study and sheds light on potential future research lines.

2 Related work

Protecting smart grid networks from malicious attacks gained much consideration in the last two decades [25–29]. For instance, in [30], Matoušek et al. investigated the utility of ICS flow monitoring in dealing with internal and external attacks against smart grids. First, the authors proposed a new ICS flow using extended IPFIX flows with application layer headers. After that, the constructed flows are used with statistical and deterministic probabilistic automata (DPA) techniques for anomaly-based attack detection. The proposed techniques were validated using IEC 60870-5-104 standard when considering rogue devices, anomalous traffics, unauthorized data downloads, and scanning attacks. Results had shown promising outcomes that depend on some tuning parameters of DPA. In [31], Jarmakiewicz et al. developed a cyber-security analysis system for a domestic power grid environment. The system correlates all information gathered from control subsystems (e.g., SCADA, IDS, and firewalls) to detect attacks and enables consistently necessary information for identifying eventual threats. Tests demonstrated that the system could detect attacks and anomalous behaviors in the grid. The general performances of such systems mainly rely on the deployed technologies and their detection capabilities. In [32], Hong et al. proposed two anomaly-based systems to tackle attacks on substations in power grids. The first system is host-based, which aims to determine the type and number of attacks affecting the substations according to several anomaly indicators (e.g., intrusions on the user interface or Intelligent Electronic Devices (IEDs), file system modifications, and IED settings alteration). The system assumed that all these indicators were available from other security mechanisms. The second system is a network-based system implemented to detect multicast message attacks, namely Generic Object Oriented Substation Event (GOOSE) and Sampled Measured Value (SMV). After being filtered, the traffic characteristics in terms of the number of messages per observation time, sequence and state number, timestamp, and data integrity are used to reveal anomalies and, eventually, attacks exploiting both messages. Besides the definition of the decision rules about these parameters that can be a challenging task in practice, they are commonly compromised by attackers.

The authors in [33] introduced pattern and flow-driven anomaly detection for communication pattern anomaly detection. Specifically, the pattern-based approach is employed to detect anomalies in communication patterns among hosts, and the flow-based approach for monitoring traffic patterns for individual flows. Results based on normal traffic reveal the promising performance of these detection schemes by reaching lower false alarms than general enterprise systems. In addition, results indicate the capacity of these schemes in detecting some common attacks launched on the MODBUS servers, including DoS attacks, scanning, system degradation, modified data, and DoS attack. In [34], Yans et al. designed an intrusion detection system by combining signature-based and model-based techniques for SCADA systems that support IEC 60870-5-104 standard. Specifically, the signature-based technique is employed to detect known attacks, and the detection signatures are defined accordingly. The model-based detector is applied to predict unknown attacks. To identify attack types, cause of transmission, and length fields are used to build a protocol-based model, and TCP connection requests, server's port numbers, and legitimate users are used to build traffic-based models. However, this IDS can detect a limited number of attacks. In [35], Lin et al. focused on the intra-SCADA exchanges and how they can be utilized to uncover related anomalies. Inter-arrival times are generated and clustered in categorical classes to construct traffic models. Then the Probabilistic suffix tree (PST) is applied to find out latent models. In [36], Kleinmann et al. introduced an attack detection scheme based on statechart. The procedure starts by converting traffic data to signal and discerns peaks and the corresponding cyclic model. Then, individual DFA (Deterministic finite automata) are created per cyclic model. Finally, the resulting DFA is used to construct the system statechart, which will be involved in attack detection. In [37], an innovative approach using a private blockchain system to detect anomalies in a smart grid network and a novel Linear Support Vector Machine Anomaly Detection (LSVMAD) approach in a fog computing (FC) environment is proposed. The LSVMAD approach achieved a detection accuracy of 89% in the FC environment and 78% in the cloud.

3 Methodology

This section presents the materials needed to design the proposed semi-supervised detection scchemes.

3.1 Deep Recurrent Autoencoder

Traditional autoencoders were designed especially for dimentionality reduction applied successfully to images, where the training approach is unsupervised without labeling data [38–40]. The key concept of the AE is replicating its input at its output. An AE comprises an encoder and a decoder, which can have multiple layers (Figure 2). It is trained in an unsupervised manner without using labeled data. The training stage aims to optimize a cost function that quantifies the deviation between the input X and its reconstruction at the output $\hat{X}$. For more details on AEs, see [41, 39].

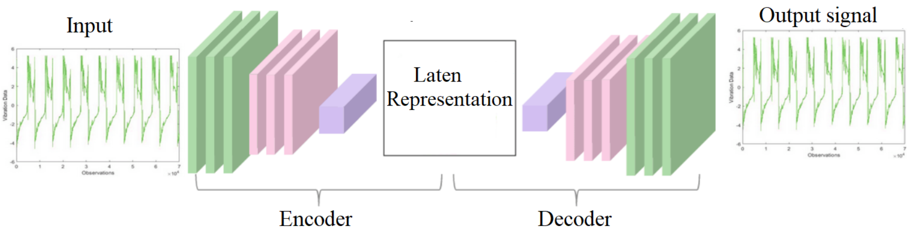

Fig. 2: Basic illustration of an AE model.

The existing recurrent model learning, such as Recurrent Neural Network (RNN) [42], Long short-term memory (LSTM) [42, 43], Gated recurrent units (GRUs) [44] are built-in supervised way. They are dedicated to capturing the time dependencies in a time series (or data sequence), with a memory cell and gates mechanism that help to remember historical data features. In this first proposed approach, we merge the desirable features of the AE and RNN-based models in one hybrid architecture to improve the encoding of time-series data in an unsupervised manner. To this end, the minimization of the reconstruction error measured using Binary Cross-Entropy (BCE) is expressed as follow:

$$\mathcal{L}_{AE-GRU} = -(X \log(\hat{X}) + (1 - X) \log(1 - \hat{X})). \tag{1}$$

The loss function represented by the BCE considers the input as binary or probability data distribution. The autoencoders are composed of two parts encoder and decoder; we incorporate two stacked Gated Recurrent Unit (GRU) into the encoder and the decoder (Figure 3). The main objective of integrating GRU is to serve as a feature extractor for time series (normal traffic) by training to reconstruct efficiently the normal data input in an unsupervised manner.

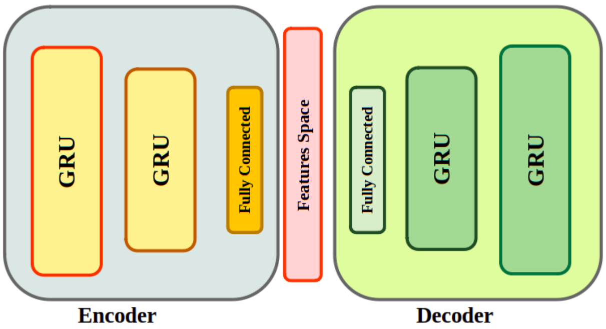

Fig. 3: The AE-GRU structure.

3.2 The hybrid GAN-GRU deep learning method

The second proposed approach is based on GAN model, which has recently emerged as an effective and efficient deep learning model for data generation and learning data representations from unlabeled data [45–47]. GAN has been successfully applied in various areas, such as image data generation and learning and time-series prediction [48]. Conventionally, GANs contain two neural networks called a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$, which are placed in an adversarial way and make GANs very flexible. They are trained in an unsupervised way, making them very attractive as labeling is an expensive task. Moreover, the GAN's generator and discriminator could be trained via only backpropagation. GAN adopts a clever procedure in training: the generator model is trained to continually generate fake data, while the discriminator model seeks to identify between true and fake (generated) data (Figure 4). In the discriminative model,

$\mathcal{D}$, learning is based on two sources of data: the training dataset (true data) and noisy data (simulated) produced by the generative model $\mathcal{G}$. After the training process, the discriminator will be able to distinguish between true and simulated data.

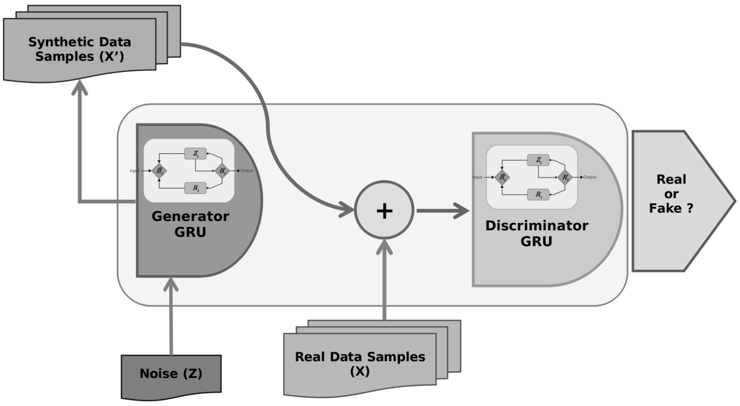

Fig. 4: The GAN-GRU structure.

GAN training optimize a cost function $\mathcal{V}(\mathcal{G}, \mathcal{D})$ as:

$$\min_{G} \max_{D} \mathcal{V}(\mathcal{G}, \mathcal{D}) = E\, p_{data(\mathbf{x})} log \mathcal{D}(\mathbf{x}) + E\, p_{g(\mathbf{x})} log(1 - \mathcal{D}(\mathcal{G}(\mathbf{z}))) \tag{2}$$

where $p_z(z)$ describe a prior on input noise variables and $p_{data}(x)$ represents the probability distribution of the true data. The distribution of samples is denoted by $p_g(x)$ produced by the generator. During the GAN training, the generator is encouraged to generate a distribution $p_g(x)$ similar to $p_{data}(x)$ of the real data. Indeed the generator helps the discriminator classify new data points (true, generated). After completing the training, the $p_g(x)$ becomes similar to $p_{data}(x)$ of the real data, which is the distribution of the historical data learned during the training by generator and discriminator.

In the GAN model, the generator and discriminator models can be any kind of neural networks, such as recurrent networks (e.g., RNN, LSTM, or GRU), making the GAN model very flexible. For instance, in the GAN-RNN, the generative and discriminative models are recurrent neural networks hence the name GAN-RNN. Indeed, the GANs are not designed for time-series modeling; however, their ability for the data distribution approximation makes them able to predict the next values of a given data sequence and thus can be investigated for improving feature extraction. The central idea of the second approach is to adopt GRUs and arrange them in an adversarial way in one architecture denoted by Generative is G-GRU, and the discriminative is D-GRU (Figure 4).

The training procedure of the proposed GAN-GRU approach remains similar to the traditional GAN model.

The proposed deep learning models will be used to model ICS traffic flow without attacks. Stable and periodic features characterize ICS communications, which is not the case with the common Internet traffic [49]. After that, the trained models can be used in anomaly detection methods to detect potential attacks in new arrival data.

3.3 Semi-supervised anomaly detection methods

Four anomaly detection methods have been adopted in our study, namely: One-Class SVM (1SVM), Local outlier factor (LO-Factor), Isolation forest (i-Forest), and Elliptic Envelope (E-Envelope). Those methods are constructed using a semi-supervised training procedure, meaning that only normal data (traffic without attack) are used in training, while the testing could include both normal and abnormal data without labeling (in the presence of attacks). The detection stage aims to distinguish between normal and abnormal traffic communications on a smart grid.

The 1SVM algorithm is one of the most popular anomaly detection techniques, known for its insensitivity to noise measurements and outliers in training. Crucially, the 1SVM is based on two essential concepts, maximizing the margin and mapping the data to a high dimensional feature space induced by a kernel function [50]. It should be noted that 1SVM is a semi-supervised binary classifier [38, 51]. More specifically, the 1SVM is constructed using unlabeled training data that contains inliers samples (anomaly-free data). In the training stage, the 1SVM process consists of estimating a boundary area, which contains most of the training data. This is carried out by determining a hyperplane with the largest distance to the nearest training data [52]. After that, the designed 1SVM is used to identify anomalies (outliers) by checking if a new test data falls within this boundary or not. Of course, testing data points are declared normal (inlier) if they are within the previously defined boundary; otherwise, they are identified as an anomaly (outliers). The 1SVM procedure assures finding a hyperplane that produces a good data separation by using kernel tricks.

The Isolation Forest approach was primarily designed by Lui in 2008 [53] and improved later in 2011 [54] to deal with anomaly detection problems where only normal observations are available. Importantly, it is an unsupervised anomaly detection approach since it is designed without the need for labeled data. The essence of the approach is founded on the principle of the Decision Tree algorithm, and it identifies anomalies by isolating outliers from the data [54]. The iF is based on the well-known Random Forest, which consists of a set (ensemble) of decision trees constructed during the training phase [55]. Isolation Forest can be considered an ensemble learning approach to deal with classification and regression problems.

In this study, we considered two other commonly used anomaly detection schemes, namely Elliptical Envelope (EE) [56] , and Local Outlier Factor (LOF) [57]. In the LOF detector, an anomaly score is computed for each observation by measuring the local divergence of the density of a given sample compared to its neighbors. In this study, the number of neighbors used in LOF is 35. In the EE detector, which aims to fit an ellipse around the data using minimum covariance determinant (MCD), the proportion of points to be included in the support of the raw MCD estimate is 0.9.

In this chapter, we will compare the performance of the AE-GRU and GAN-GRU-based 1SVM, LOF, i-Forest, and EE methods for cyber-attack detection in a smart grid. Next, we will discuss how these two models (i.e., AE-GRU and GAN-GRU) can be used for cyber-attack detection.

3.4 The proposed deep-learning-driven anomaly detection framework

This study presents semi-supervised hybrid deep learning methods to detect cyber attacks in smart grids, particularly monitoring smart grid communication networks. The proposed approach constitutes a framework for online monitoring of the communications (traffic) within an industrial network based on IEC-60780-5-104 (aka IEC 104), often used for smart grids' substation control [58]. Importantly, the IEC-104 protocol, widely adopted in Europe, represents an international standard of data transmission between a power SCADA center and outstations via TCP/IP [59]. This protocol enables connecting MTUs and RTUs using a standard TCP/IP network. Specifically, the IEC-104 protocol ensures data transmission in two directions: from an RTU to the MTU and vice-versa. More details about IEC-104 protocol can be found in [35, 58].

Generally speaking, network communication consists of timed packets sent. Some of them wait for confirmation and have a cause of transmission. Normal communication can be seen as data sequences and time-series data. Furthermore, temporal data dependencies have to be modeled to learn the data distribution of normal traffic. In this chapter, the designed deep learning models will be used to capture the evolution in normal data. The AE-GRU and GAN-GRU models integrate GRU models in their structure, which are powerful deep learning models designed with a gating mechanism and memory cell, allowing them to learn long-range dependencies. As discussed above, for instance, in the AE-GRU model, the autoencoder performs two essential tasks I) features extraction and II) dimensionality reduction. Moreover, deep AE constructs a new compact representation that incorporates pertinent features that we call features space built during the training procedure. We combined the robustness of the deep autoencoders and the effectiveness of the recurrent neural network in order to design a deep recurrent autoencoder model, able to learn learning lengthy-time period dependencies. The proposed framework is a data-driven approach based on learning from data. The proposed hybrid model learns normal traffic, where the communication responds to the IEC-104 protocol constraints (request, response, acknowledgment).

Figure 5 illustrates the flowchart of the proposed cyber-attack detection procedure. The experiment is divided into the training phase and the detection phase. In more detail, we remove missing values and standardize numeric features in the data preprocessing step. In the training step, we train the neural network and anomaly detection procedures using attack-free data. Finally, the trained model is verified using the testing data.

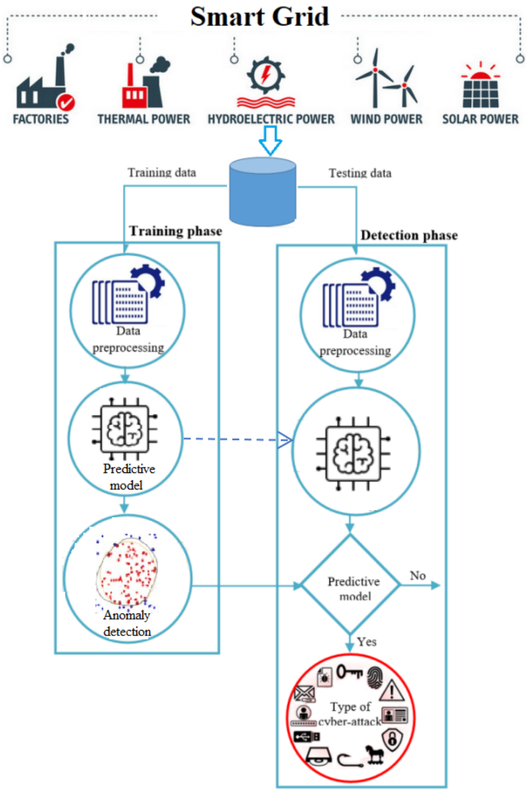

Fig. 5: Flowchart of the detection framework.

In the training stage, the data is preprocessed by normalization and then arranging data into a sequence with a given length. The next step consists of learning to encode the data sequence into a compact representation that contains pertinent features. Furthermore, the training procedure aims to minimize the reconstruction error of the encoded data sequence; this step is repeated until the convergence of the trained model and until the stabilization of the reconstruction error. In fact, the result of this step is a feature space of normal traffic communications with no attack, which is used to feed the anomaly detection considered methods in this study. Indeed, the anomaly detection considered methods are trained using a compact version of the normal data and tested later with data containing attacks and normal communication to evaluate the detection performance.

Here, we evaluate two types of autoencoders, the AE-GRU and the GAN-GRU. Recall here that the training of the AE-GRU is accomplished by minimizing the reconstruction error of the training dataset using the cross-entropy loss function. In the GAN-GRU, the discriminator is trained to figure out how to distinguish between real data (training dataset) from fake generated by the generator using noisy data. In other words, the discriminator learns to recognize the generator's flaws.

The anomaly detection approaches used are trained based on the resulting feature space containing the encoding of normal data. Furthermore, when a data sequence belongs to a given attack category, its numerical signature is encoded with model parameters learned using normal data, it will be certainly different from normal encoding, which can be sensed by the detectors.

In this study, five statistical scores are employed to quantify the performance of the studied methods computed using a 2 × 2 confusion matrix: Accuracy, Precision, Recall, F1-score, and Area under curve (AUC). For a binary detection problem, the number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) are used to compute the evaluation metrics.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{3}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{5}$$

$$\text{F1} - \text{score} = 2\frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{6}$$

4 Results and discussion

We conducted several experiments to evaluate the performance of the considered anomaly detection models used to design the intrusion detection framework. Indeed, we assess the intrusion detection system performance with several attacks, namely: one man-in-the-middle (MITM) attack, DOS, Connection loss, injection, spoofing, scanning, switching, replaying RTU, blocking HMI, replaying HMI, value changing, masquerading and Rouge devices. The present study adopted two deep recurrent auto-encoders combined with four anomaly detection methods to design a data-driven intrusion detection framework. In order to validate the effectiveness of the proposed approach, we used three datasets of smart grid ICS containing normal and various attacks on IEC 104 communications.

4.1 Data description

This part is devoted to assessing the efficiency of the proposed deep learning-driven approaches in detecting cyber-attacks in the ICS traffic flow of smart grids. The study is accomplished through actual data from a publicly available database provided in [60]. Three different datasets comprise normal traffic data, and different types of attacks are considered in this study. The first dataset, called BUT-IEC104-I, was generated by from Brno University of technology [30, 61] and contains traffic records from a real smart grid that supports the industrial network standards IEC 608705-104. The second data, called VRT-IEC 104, represent various attacks on IEC 104 communication created using IEC virtual testbed developed at Brno University of Technology. The dataset was captured at the HMI side in the topology. The third data, called GICS-MMS, comprises a set of cyber security attacks on MMS communication that were created manually on the real-life communication provided by the G-ICS labs, University of Grenoble Alps, France (http://lig-g-ics.imag.fr). More details about the used dataset can be found in [60]

Next, the developed AE-GRU-based andGAN-GRU-based 1SVM, iF, LOF, and EE methods will be assessed using three diffrent datasets.

Case 1): cyber-attack detection in BUT-IEC104-I dataset The dataset was created by Matousek et al. [30, 61] from Brno University of technology. It provides traffic records from an actual smart grid that supports the industrial network standards IEC 608705-104, commonly known as IEC 104 and IEC 61850. Traffic monitoring has been carried out using the IPFIX traffic monitor. The captures consist of IPFIX flow added to application protocols headers. The resulting data includes several traffic features: IP addresses, ports, object id, and other derived characteristics (e.g., start and end times and quantity of exchanged data). Besides the normal traffic, this dataset contains the following scenarios with attacks.

- **DOS attack against an IEC 104 control station:** DOS attack has intended to crash a control station and collapse the grid accordingly. The attacker gets access using a spoofed IP address and floods the victim with 1049 messages in 30 minutes.
- **Switching attack:** A malware-based attack consists of switching on/off the targeted devices. Within this attack, 72 packets were sent in an interval time of 10 minutes.
- **Injection commands attacks:** in this case, the attacker manipulates a connected device, then changes its settings or inserts false commands to generate unusual actions and eventually creates different sorts of anomalies. Two scenarios of 5 and 15 minutes in which 83 and 221 packets were injected, respectively.
- **Connection loss attacks:** In this case, the attacker tries to disconnect targeted devices and break down the attached communications. Two scenarios

were implemented. In the first one, connection loss during 10 minutes resulted in 146 missing packets. In the second scenario, connection loss for one hour, causing losses of packets.

- **Rogue devices attack:** this attack takes place when unauthorized devices are attributed to the communication network, allowing them to share random messages with legitimate devices, which can also cause in random responses and unpredicted actions, which are generally destructive. The dataset includes an attack of 30 minutes, generating 417 messages.

At first, the models were trained based on training data. The training data contains normal IEC 104 communication without attacks; it contains 58930 packets recorded in 2 days and 19:55 hours of traffic. At each time point, the extracted from the IEC 104 packet contains 16 features of traffic variables, including source IP address, destination IP address, source port, destination port, and destination port. The values of the tuned parameters of the trained models are listed in Table 1. All the hyper-parameters are computed during the models training by the minimization of the cross-entropy of the reconstructed error.

Table 1: Hyper-parameters used for the experiment.

Model	parameter	value
	Input features	9
	activation function	ReLu
	loss function	Cross Entropy
	Optimizer	Rmsprop
	Epoch	300
	Batch size	250
	Timesteps	12
AE-GRU & GAN-RNN		
Encoder	Layers	3
	layer1	GRU(units=128)
	layer2	GRU(units=16)
	layer3	Dense(units=Features)
Decoder	layers	
	layer1	Dense(units=Features)
	layer2	GRU(units=16)
	layer3	GRU(units=128)
iF	contamination	0.01
	estimators	150
LOF	contamination	0.01
	novelty	TRUE
EE	contamination	0.01
	support_fraction	0.995
1SVM	kernel	RBF
	ν	0.0015
	γ	0.25

Detection results of the proposed AE-GRU and GAN-GRU-based intrusion detection methods when applied to the but-iec104-i dataset are listed in Tables 2 and 3, respectively. The adopted anomaly detection approaches are trained using the features space of normal data, and testing data contain both normal traffic with different attacks. For the DOS attack, Table 2 indicates that the AE-GRU-based EE method recorded the best performance, while the AE-GRU-based 1SVM comes in the second place with an F1-score of 0.9968, followed by the AE-GRU-iForest with an F1-score of 0.9052. The AE-GRU-LOF scores the best detection score for the Switching attack with an F1-score of 0.942, followed by the AE-GRU-iForest with an F1-score of 0.9404. Here, the AE-GRU-EE detection performance reaches an F1-score of 0.9281. Results in Table 2 show that the injection attack has been detected by all adopted anomaly detection methods with an F1-score around 0.98. The AE-GRU-EE method dominates the other models for detecting connection loss attacks with an F1-score of 0.9643. AE-GRU-based LOF and iForest follow it by reaching, respectively, F1-score values of 0.9493 and 0.9493. Moreover, the AE-GRU-1SVM scheme provides the best detection for Rogue Devices attacks with an F1-score of 0.9968, while the AE-GRU-based iForest and LOF schemes provide comparable performance with an F1-score of 0.9903. Of course, we noticed that the AE-GRU-1SVM scheme is providing poor results for both switching and connection loss attacks even after tuning its hyper-parameters ν and γ, which are very sensitive and impact the

Table 2: AE-GRU-based anomaly detection schemes using but-iec104-i dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
DOS	iForest	0.8875	1	0.8269	0.9052
DOS	LOF	0.65	0.65	1	0.7879
DOS	1SVM	0.9958	1	0.9936	0.9968
DOS	EE	1	1	1	1
Switch	iForest	0.8903	0.9856	0.8991	0.9404
Switch	LOF	0.8945	1	0.8904	0.942
Switch	1SVM	0.1181	1	0.0833	0.1538
Switch	EE	0.8692	0.9852	0.8772	0.9281
Injection	iForest	0.9619	1	0.9616	0.9804
Injection	LOF	0.9619	1	0.9616	0.9804
Injection	1SVM	0.9771	1	0.977	0.9884
Injection	EE	0.9619	1	0.9616	0.9804
ConnLoss	iForest	0.8017	0.9789	0.8158	0.8899
ConnLoss	LOF	0.9052	1	0.9035	0.9493
ConnLoss	1SVM	0.3103	1	0.2982	0.4594
ConnLoss	EE	0.931	0.9818	0.9474	0.9643
Rogue Dev	iForest	0.9853	1	0.9808	0.9903
Rogue Dev	LOF	0.9853	1	0.9808	0.9903
Rogue Dev	1SVM	0.9951	1	0.9936	0.9968
Rogue Dev	EE	0.7647	0.7647	1	0.8667

detection, in contrast to the other methods where the most important parameter is the contamination that indicates the percentage of abnormal data (outliers) used during the training.

From Table 3, there is not a single model that is uniformly superior to others. We observe that GAN-GRU-based iForest and LOF schemes reach satisfactory detection performance for the five considered attacks. For instance, the GAN-GRU-LOF scheme obtained the best performance for DOS attack detection with 0.9936, followed by the GAN-GRU-EE scheme with an F1-score of 0.987 and *One-Class SVM* with 0.9315. It is worth noting that combining the proposed hybrid models with anomaly detection methods provides a promising tool in detecting attacks in ICS traffic of smart grids.

Table 3: GAN-GRU adopted anomaly detection schemes using but-iec104-i dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
DOS	iForest	0.8	1	0.6923	0.8182
DOS	LOF	0.9917	1	0.9872	0.9936
DOS	1SVM	0.9167	1	0.8718	0.9315
DOS	EE	0.9833	1	0.9744	0.987
Switch	iForest	0.8439	0.9848	0.8509	0.913
Switch	LOF	0.9873	1	0.9868	0.9934
Switch	1SVM	0.1814	1	0.1491	0.2595
Switch	EE	0.9789	0.9869	0.9912	0.989
Injection	iForest	0.9553	0.9932	0.9616	0.9771
Injection	LOF	0.9891	1	0.989	0.9945
Injection	1SVM	0.0229	1	0.0164	0.0323
Injection	EE	0.9826	1	0.9825	0.9912
ConnLoss	iForest	0.7069	1	0.7018	0.8248
ConnLoss	LOF	0.9052	1	0.9035	0.9493
ConnLoss	1SVM	0.8194	1	0.8158	0.8986
ConnLoss	EE	0.9677	0.9804	0.9868	0.9836
Rogue Dev	iForest	0.9853	1	0.9808	0.9903
Rogue Dev	LOF	0.7549	0.7624	0.9872	0.8604
Rogue Dev	1SVM	0.25	1	0.0192	0.0377
Rogue Dev	EE	0.7549	0.7624	0.9872	0.8604

To conclude the assessment of the proposed AE-GRU and GAN-GRU-based intrusion detection schemes when applied to the but-iec104-i dataset, Table 4 lists the averaged evaluation scores for each scheme. Results show that combining the proposed hybrid deep learning models with anomaly detection methods provides a promising tool for monitoring ICS traffic flow without using labeled data. It can be seen that the GAN-GRU-based LOF and EE schemes dominate the other models by reaching an F-score of 0.96. They are followed by the AE-GRU-based EE scheme with an F-score of 0.96.

Table 4: Avreaged performance of AE-GRU and GAN-GRU-based methods using but-iec104-i dataset.

Method	Accuracy	Precision	Recall	F1-score
AE-GRU-iForest	0.91	0.99	0.90	0.94
AE-GRU-LOF	0.90	0.94	0.95	0.94
AE-GRU-1SVM	0.68	1.00	0.67	0.72
AE-GRU-EE	0.91	0.95	0.96	0.95
GAN-GRU-iForest	0.86	1.00	0.84	0.90
GAN-GRU-LOF	0.93	0.95	0.97	0.96
GAN-GRU-1SVM	0.44	1.00	0.37	0.43
GAN-GRU-EE	0.93	0.95	0.98	0.96

Case 2): cyber-attack detection in G-ICS dataset The second experiment evaluates the proposed cyber-attack detection schemes using the G-ICS created University of Grenoble Alps, France. The G-ICS data contains a set of cyber security attacks on MMS communication, including lost connection, injection, scanning, and interrupt attacks. In addition, this data comprises normal communication data, which is used for training the proposed models.

The data contains two lost connections, with 58 missing packets in the first and 63 missing packets in the second. The data include scanning attacks, also known as reconnaissance attacks, which are performed to obtain necessary information about network topology and components. Here, the attack was launched at 07:50:41,51 and ended at 07:53:01:51. In the G-ICS, We find also interrupt attacks, where an attacker tries to interrupt MMS services, and modification attacks, where an attacker tries modifying packets.

Tables 5 and 6 shows detection results obtained by AE-GRU and GAN-GRU-based anomaly detection schmes using G-ICS dataset. As can be seen in Table 5, for the modification attacks, the AE-GRU-1SVM achieved the best performance, followed by the AE-GRU-iForest with an F-score of 0.9735 and the AE-GRU-EE with an F-score of 0.9547, while the AE-GRU-LOF recorded an F1-score of 0.9167. The AE-GRU-EE recorded the best performance for the connection loss attack with an F1-score of 0.9915, followed by the AE-GRU-based iForest and LOF schemes with an F1-score of 0.9783, and 0.9646, respectively. In contrast, the AE-GRU-1SVM did not record a good performance, with an F1-score of 0.7579. However, for scan resources attack, all adopted methods have recorded a high F1-score greater than 0.97, except the AE-GRU-EE with 0.9204. For injection attack detection, AE-GRU-EE dominates the other detectors. It is followed by the AE-GRU-based 1SVM, LOF, and iForest schemes with F1-score values of 0.9839, 0.9043, and 0.8333, respectively. Similar conclusions hold true also when using the GAN-GRU-based anomaly detection schemes (Table 6). There is no single approach dominating all models for the considered attacks.

As shown above, no unique approach is uniformly superior to others in detecting different attacks. Table 7 presents the aggregated performances of each

Table 5: AE-GRU adopted anomaly detection approaches using GICS dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
Modification	iForest	0.9538	1	0.9483	0.9735
Modification	LOF	0.8462	0.8871	0.9483	0.9167
Modification	1SVM	1	1	1	1
Modification	EE	0.9154	0.9134	1	0.9547
ConnLoss	iForest	0.9576	0.9912	0.9658	0.9783
ConnLoss	LOF	0.9322	1	0.9316	0.9646
ConnLoss	1SVM	0.6102	0.9863	0.6154	0.7579
ConnLoss	EE	0.9831	0.9915	0.9915	0.9915
scanning	iForest	0.9524	1	0.9483	0.9735
scanning	LOF	0.9524	0.9508	1	0.9748
scanning	1SVM	0.9524	0.9508	1	0.9748
scanning	EE	0.8571	0.9455	0.8966	0.9204
Injection	iForest	0.7273	1	0.7143	0.8333
Injection	LOF	0.8333	1	0.8254	0.9043
Injection	1SVM	0.9697	1	0.9683	0.9839
Injection	EE	1	1	1	1

Table 6: GAN-RNN adopted anomaly detection approaches using GICS dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
ConnLoss	iForest	0.6695	0.9875	0.6752	0.802
ConnLoss	LOF	0.8898	1	0.8889	0.9412
ConnLoss	1SVM	0.7881	1	0.7863	0.8804
ConnLoss	EE	0.9831	0.9915	0.9915	0.9915
Injection	iForest	0.8485	1	0.8413	0.9138
Injection	LOF	0.9545	0.9545	1	0.9767
Injection	1SVM	0.9545	1	0.9524	0.9756
Injection	EE	1	1	1	1
scanning	iForest	0.9206	1	0.9138	0.955
scanning	LOF	0.9048	0.9483	0.9483	0.9483
scanning	1SVM	0.9841	0.9831	1	0.9915
scanning	EE	0.9365	0.9355	1	0.9667
Modification	iForest	0.8231	1	0.8017	0.8899
Modification	LOF	0.8769	1	0.8621	0.9259
Modification	1SVM	0.9692	0.9667	1	0.9831
Modification	EE	0.9769	0.9748	1	0.9872

cyber-attack detection scheme. In terms of all evaluation scores computed, the GAN-GRU-EE scheme is the best approach to detect attacks that occurred in the G-ICS dataset with high efficiency by obtaining an averaged F1-score of 0.99. It is followed by the AE-GRU-EE scheme with an averaged F1-score of 0.97. It could be due to the capability of the hybrid deep learning models (i.e., AE-GRU and GAN-GRU) to extract relevant features from ICS traffic data and the sensitivity of the anomaly detection method. Overall, we should highlight the potential of merging hybrid deep learning models with anomaly detection methods in detecting cyber-attacks in ICS traffic flow.

Table 7: Avreaged performance of AE-GRU and GAN-GRU-based methods using G-ICS dataset.

Method	Accuracy	Precision	Recall	F1-score
AE-GRU-iForest	0.90	1.00	0.89	0.94
AE-GRU-LOF	0.89	0.96	0.93	0.94
AE-GRU-1SVM	0.88	0.98	0.90	0.93
AE-GRU-EE	0.94	0.96	0.97	0.97
GAN-GRU-iForest	0.82	1.00	0.81	0.89
GAN-GRU-LOF	0.91	0.98	0.92	0.95
GAN-GRU-1SVM	0.92	0.99	0.93	0.96
GAN-GRU-EE	0.97	0.98	1.00	0.99

These results prove that the amalgamation of the hybrid deep learning models with anomaly detection methods provided satisfactory results in identifying cyber-attacks in ICS traffic flow.

Case 3): cyber-attack detection in vrt-iec104 dataset The final experiment aims to assess the potential of the proposed technique in detecting various attacks on IEC 104 communication created using the IEC virtual testbed developed by Peter Grofcik at Brno University of Technology [60]. Since IEC 104 packets may include multiple APDUs (application protocol data units) and ASDUs (application service data units), the IPFIX probe separates these packets into so-called virtual IEC 104 packets used for further analysis. The dataset was captured at the HMI side in the topology [60]. The propose models have been trained using three days of normal communication data with no attacks (i.e., total of 381.666 virtual packets). Then, we assessed the trained models using data with different types of attacks, including report-block-HMI attacks, HMI-MITM attacks, reply-HMI attacks, value-change-HMI attacks,and masquerating attacks. More details about the considered attacks can be find in the readme file in [60].

Table 8 reports the detection results of the AE-GRU-based anomaly detection methods when applied to the vrt-iec104 dataset. Results indicate the AE-GRU-LOF recorded the highest detection accuracy for detecting Command Act attacks, with an F1-score of 0.9831; it is followed by the AE-GRU-EE scheme with an F1-score of 0.9565. For the other attacks, we observe that the AE-GRU-based EE and LOF schemes recorded the highest F1-score greater than 0.95, while the other methods did perform well.

Table 9 lists detection results of the GAN-GRU-based anomaly detection schemes using the vrt-iec104 dataset. We observe that GAN-GRU-based iForest

Table 8: AE-GRU adopted anomaly detection approaches using vrt-iec104 dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
command Act	iForest	0.7396	0.7188	0.9583	0.8214
command Act	LOF	0.9792	1	0.9667	0.9831
command Act	1SVM	0.5729	0.6092	0.8833	0.7211
command Act	EE	0.9479	1	0.9167	0.9565
Modification	iForest	0.1282	0.1239	0.975	0.2199
Modification	LOF	0.9958	1	0.9667	0.9831
Modification	1SVM	0.0987	0.1006	0.775	0.1781
Modification	EE	0.9958	1	0.9667	0.9831
block-HMI	iForest	0.4579	0.124	0.95	0.2194
block-HMI	LOF	0.9973	1	0.9667	0.9831
block-HMI	1SVM	0.0608	0.0614	0.75	0.1135
block-HMI	EE	0.9973	1	0.9667	0.9831
replaying HMI	iForest	0.2989	0.0933	0.9417	0.1698
replaying HMI	LOF	0.9975	1	0.9667	0.9831
replaying HMI	1SVM	0.0647	0.0638	0.825	0.1184
replaying HMI	EE	0.9937	1	0.9167	0.9565
replaying RTU	iForest	0.3438	0.2452	0.95	0.3898
replaying RTU	LOF	0.9926	1	0.9667	0.9831
replaying RTU	1SVM	0.3768	0.2459	0.8833	0.3847
replaying RTU	EE	0.9798	1	0.9083	0.9519
change of HMI	iForest	0.1571	0.1064	0.9583	0.1915
change of HMI	LOF	0.9965	1	0.9667	0.9831
change of HMI	1SVM	0.2335	0.0928	0.725	0.1645
change of HMI	EE	0.9913	1	0.9167	0.9565
change of RTU	iForest	0.2478	0.0997	0.95	0.1805
change of RTU	LOF	0.9971	1	0.9667	0.9831
change of RTU	1SVM	0.077	0.0778	0.8833	0.143
change of RTU	EE	0.9964	1	0.9583	0.9787
masquerating	iForest	0.3289	0.0497	0.9417	0.0944
masquerating	LOF	0.9985	1	0.9583	0.9787
masquerating	1SVM	0.2067	0.0282	0.6083	0.0539
masquerating	EE	0.9966	1	0.9083	0.9519

and LOF schemes recorded the highest F1-score of 9831 compared to GAN-GRU-based 1SVM and EE schemes that achieved an under 0.78 which is not satisfactory. For the Changes value and replaying HMI attacks, the GAN-GRU-LOF was the only method that satisfactory detected this kind of attack with an F1-score of 0.9831. The GAN-GRU-LOF reached the best performance with an F1-score of 0.9831 regarding the attack replaying RTU; after that comes the GAN-GRU-1SVM with an F1-score of 0.806. In contrast, the other methods could not detect this kind of attack properly. The same performance detection was recorded for the attacks: masquerading, change of HMI, and change of RTU, where the GAN-GRU-LOF scheme is the only able to detect the attacks with an F1-score of 0.9831.

Table 9: GAN-RNN adopted anomaly detection approaches using vrt-iec104 dataset.

Attack	Method	Accuracy	Precision	Recall	F1-score
command Act	iForest	0.9792	1	0.9667	0.9831
command Act	LOF	0.9792	1	0.9667	0.9831
command Act	1SVM	0.7188	0.7845	0.7583	0.7712
command Act	EE	0.5885	0.6108	0.9417	0.741
changes Value	iForest	0.3792	0.1612	0.9333	0.2749
changes Value	LOF	0.9958	1	0.9667	0.9831
changes Value	1SVM	0.6029	0.1943	0.6833	0.3026
changes Value	EE	0.1786	0.1247	0.9167	0.2195
block-HMI	iForest	0.996	1	0.95	0.9744
block-HMI	LOF	0.9973	1	0.9667	0.9831
block-HMI	1SVM	0.9646	1	0.5583	0.7166
block-HMI	EE	0.9967	1	0.9583	0.9787
replaying HMI	iForest	0.5615	0.1418	0.9417	0.2465
replaying HMI	LOF	0.9975	1	0.9667	0.9831
replaying HMI	1SVM	0.8598	0.2735	0.5083	0.3556
replaying HMI	EE	0.1523	0.0766	0.9167	0.1414
replaying RTU	iForest	0.3989	0.2653	0.975	0.4171
replaying RTU	LOF	0.9926	1	0.9667	0.9831
replaying RTU	1SVM	0.9283	1	0.675	0.806
replaying RTU	EE	0.204	0.2075	0.925	0.339
change of HMI	iForest	0.2101	0.1143	0.975	0.2046
change of HMI	LOF	0.9965	1	0.9667	0.9831
change of HMI	1SVM	0.9158	0.5804	0.6917	0.6312
change of HMI	EE	0.3168	0.129	0.9667	0.2276
change of RTU	iForest	0.96	1	0.5417	0.7027
change of RTU	LOF	0.9971	1	0.9667	0.9831
change of RTU	1SVM	0.915	0.5101	0.6333	0.5651
change of RTU	EE	0.2515	0.1016	0.9667	0.1839
masquerating	iForest	0.9817	1	0.5083	0.674
masquerating	LOF	0.9988	1	0.9667	0.9831
masquerating	1SVM	0.9202	0.1835	0.3333	0.2367
masquerating	EE	0.1064	0.0387	0.9667	0.0744

From results in Tables 8 and 9, the main finding is that AE-GRU and GAN-GRU-based LOF schemes have maintained a high detection performance for all attacks. Moreover, the remaining methods could not keep a high detection accept for the block of HMI.

Table 10: Avreaged performance of AE-GRU and GAN-GRU-based methods using vrt-iec104 dataset.

Method	Accuracy	Precision	Recall	F1-score
AE-GRU-iForest	0.34	0.20	0.95	0.29
AE-GRU-LOF	0.99	1.00	0.97	0.98
AE-GRU-1SVM	0.21	0.16	0.79	0.23
AE-GRU-EE	0.99	1.00	0.93	0.96
GAN-GRU-iForest	0.68	0.59	0.85	0.56
GAN-GRU-LOF	0.99	1.00	0.97	0.98
GAN-GRU-1SVM	0.85	0.57	0.61	0.55
GAN-GRU-EE	0.35	0.29	0.94	0.36

Table 11 displays the aggregated performances of each approach based on the three considered datasets. We can see that the AE-GRU and GAN-GRU-based LOF schemes achieved the best detection performance of all attacks, with an F1-score of 0.98. They dominate the other investigated deep learning-based anomaly detection methods.

Table 11: Avreaged performance of AE-GRU and GAN-GRU-based methods using the three datasets.

Method	Accuracy	Precision	Recall	F1-score
AE-GRU-iForest	0.34	0.20	0.95	0.29
AE-GRU-LOF	0.99	1.00	0.97	0.98
AE-GRU-1SVM	0.21	0.16	0.79	0.23
AE-GRU-EE	0.99	1.00	0.93	0.96
GAN-GRU-iForest	0.68	0.59	0.85	0.56
GAN-GRU-LOF	0.99	1.00	0.97	0.98
GAN-GRU-1SVM	0.85	0.57	0.61	0.55
GAN-GRU-EE	0.35	0.29	0.94	0.36

5 Conclusion

Accurately detecting attacks in ICS communication networks is undoubtedly necessary for designing modern smart grids and ensuring safe and reliable operation. This chapter introduced data-driven anomaly detection approaches for intrusion detection in a smart grid environment by monitoring ICS traffic communication. At first, we designed two-hybrid deep learning models, AE-GRU and GAN-GRU, by integrating AE and GRU's desirable characteristics with the GRU model's capacity to capture the time-dependence in ICS traffic data. Then, these two models have been combined with semi-supervised anomaly detection techniques to design semi-supervised deep recurrent-based anomaly detection schemes for ICS communication monitoring in smart grids. The AE-GRU and GAN-GRU models aim to automatically learn and capture the relevant features from ICS traffic flow, and the anomaly detection schemes (i.e., iForest, LOF, 1SVM, and EE) are applied to extracted features to detect cyber-attacks based. Three publically available datasets are used to assess the performance of the proposed approaches. Four statistical scores have been utilized to judge the detection accuracy of the studied methods, including accuracy, precision, recall, and F1-score. Results revealed that the proposed AE-GRU-based EE and LOF methods offer superior detection performance of different cyber-attack types and dominate the other investigated methods.

Despite the improved detection performance of the proposed semi-supervised deep learning-based techniques for detecting anomalies in ICS traffic on the IEC 104 communication, future works will be aimed to extend further the range of their application to monitor ICS communication of industrial systems. Further, we plan to improve the robustness of the hybrid deep learning models (AE-GRU and GAN-GRU) to noisy measurements by developing a wavelet-based hybrid deep learning detector. To this end, we will use wavelet decomposition to capture multivariate information in the time and frequency domains and then employ a hybrid deep learning model to extract relevant features that will be fed to the anomaly detection methods for cyber-attack detection. It is expected that by applying wavelet-based multiscale denoising to the received signals, noise effects will be reduced, thus improving cyber-attack detection. Another interesting direction for future work is the design of unsupervised cyber-attack detection strategy by integrating unsupervised deep learning methods as features extractors, such as deep variational auto-encoder [62], with the sensitivity of statistical monitoring charts, such as Generalized Likelihood Ratio Test [63].

Acknowledgments. This publication is based upon work supported by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800.

References

1. W. Wang and Z. Lu, "Cyber security in the smart grid: Survey and challenges," *Computer networks*, vol. 57, no. 5, pp. 1344–1371, 2013.
2. M. Z. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer networks*, vol. 169, p. 107094, 2020.
3. R. Leszczyna, "A review of standards with cybersecurity requirements for smart grid," *Computers & security*, vol. 77, pp. 262–276, 2018.

4. R. K. Pandey and M. Misra, "Cyber security threats—smart grid infrastructure," in *2016 National power systems conference (NPSC)*. IEEE, 2016, pp. 1–6.
5. A. Teymouri, A. Mehrizi-Sani, and C.-C. Liu, "Cyber security risk assessment of solar pv units with reactive power capability," in *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2018, pp. 2872–2877.
6. W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems," *Cluster Computing*, vol. 25, no. 1, pp. 561–578, 2022.
7. M. Stănculescu, S. Deleanu, P. C. Andrei, and H. Andrei, "A case study of an industrial power plant under cyberattack: Simulation and analysis," *Energies*, vol. 14, no. 9, p. 2568, 2021.
8. A. A. Z. Khan and G. Serpen, "Intrusion detection and identification system design and performance evaluation for industrial scada networks," *arXiv preprint arXiv:2012.09707*, 2020.
9. J. R. Vacca, *Cyber security and IT infrastructure protection*. Syngress, 2013.
10. M. Touhiduzzaman, S. N. G. Gourisetti, C. Eppinger, and A. Somani, "A review of cybersecurity risk and consequences for critical infrastructure," *2019 Resilience Week (RWS)*, vol. 1, pp. 7–13, 2019.
11. J. Jiang, X. Zhao, S. Wallace, E. Cotilla-Sanchez, and R. Bass, "Mining pmu data streams to improve electric power system resilience," in *Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, 2017, pp. 95–102.
12. C. Konstantinou, M. Sazos, and M. Maniatakos, "Attacking the smart grid using public information," in *2016 17th Latin-American Test Symposium (LATS)*. IEEE, 2016, pp. 105–110.
13. S. Basumallik, R. Ma, and S. Eftekharnejad, "Packet-data anomaly detection in pmu-based state estimator using convolutional neural network," *International Journal of Electrical Power & Energy Systems*, vol. 107, pp. 690–702, 2019.
14. A. Walker, J. Desai, D. Saleem, and T. Gunda, "Cybersecurity in photovoltaic plant operations," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2021.
15. J. Ye, A. Giani, A. Elasser, S. K. Mazumder, C. Farnell, H. A. Mantooth, T. Kim, J. Liu, B. Chen, G.-S. Seo *et al.*, "A review of cyber-physical security for photovoltaic systems," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2021.
16. C.-C. Sun, A. Hahn, and C.-C. Liu, "Cyber security of a power grid: State-of-the-art," *International Journal of Electrical Power & Energy Systems*, vol. 99, pp. 45–56, 2018.
17. Z. El Mrabet, N. Kaabouch, H. El Ghazi, and H. El Ghazi, "Cyber-security in smart grid: Survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018.
18. F. Nejabatkhah, Y. W. Li, H. Liang, and R. Reza Ahrabi, "Cyber-security of smart microgrids: A survey," *Energies*, vol. 14, no. 1, p. 27, 2020.
19. Y. Zhang, L. Wang, Z. Liu, and W. Wei, "A cyber-insurance scheme for water distribution systems considering malicious cyberattacks," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1855–1867, 2020.
20. A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures π," in

2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). IEEE, 2011, pp. 232–237.

21. D. An, Q. Yang, W. Liu, and Y. Zhang, “Defending against data integrity attacks in smart grid: A deep reinforcement learning-based approach,” *IEEE Access*, vol. 7, pp. 110 835–110 845, 2019.
22. P. Srikantha and D. Kundur, “Denial of service attacks and mitigation for stability in cyber-enabled power grid,” in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT).* IEEE, 2015, pp. 1–5.
23. M. Z. Gunduz and R. Das, “Analysis of cyber-attacks on smart grid applications,” in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP).* IEEE, 2018, pp. 1–5.
24. A. Huseinovic, S. Mrdovic, K. Bicakci, and S. Uludag, “A taxonomy of the emerging denial-of-service attacks in the smart grid and countermeasures,” in *2018 26th Telecommunications Forum (TELFOR).* IEEE, 2018, pp. 1–4.
25. Z. Lu, X. Lu, W. Wang, and C. Wang, “Review and evaluation of security threats on the communication networks in the smart grid,” in *2010-Milcom 2010 Military Communications Conference.* IEEE, 2010, pp. 1830–1835.
26. S. A. Yadav, S. R. Kumar, S. Sharma, and A. Singh, “A review of possibilities and solutions of cyber attacks in smart grids,” in *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH).* IEEE, 2016, pp. 60–63.
27. M. Caselli, E. Zambon, and F. Kargl, “Sequence-aware intrusion detection in industrial control systems,” in *Proceedings of the 1st ACM Workshop on Cyber-Physical System Security*, 2015, pp. 13–24.
28. H. Yoo and T. Shon, “Novel approach for detecting network anomalies for substation automation based on iec 61850,” *Multimedia Tools and Applications*, vol. 74, no. 1, pp. 303–318, 2015.
29. P. Maynard, K. McLaughlin, and B. Haberler, “Towards understanding man-in-the-middle attacks on iec 60870-5-104 scada networks,” in *2nd International Symposium for ICS & SCADA Cyber Security Research 2014 (ICS-CSR 2014) 2*, 2014, pp. 30–42.
30. P. Matoušek, O. Ryšavỳ, M. Grégr, and V. Havlena, “Flow based monitoring of ics communication in the smart grid,” *Journal of Information Security and Applications*, vol. 54, p. 102535, 2020.
31. J. Jarmakiewicz, K. Parobczak, and K. Maślanka, “Cybersecurity protection for power grid control infrastructures,” *International Journal of Critical Infrastructure Protection*, vol. 18, pp. 20–33, 2017.
32. J. Hong, C.-C. Liu, and M. Govindarasu, “Integrated anomaly detection for cyber security of the substations,” *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1643–1653, 2014.
33. A. Valdes and S. Cheung, “Communication pattern anomaly detection in process control systems,” in *2009 IEEE Conference on Technologies for Homeland Security.* IEEE, 2009, pp. 22–29.
34. Y. Yang, K. McLaughlin, T. Littler, S. Sezer, B. Pranggono, and H. Wang, “Intrusion detection system for iec 60870-5-104 based scada networks,” in *2013 IEEE power & energy society general meeting.* IEEE, 2013, pp. 1–5.
35. C.-Y. Lin and S. Nadjm-Tehrani, “Understanding iec-60870-5-104 traffic patterns in scada networks,” in *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security*, 2018, pp. 51–60.

36. A. Kleinmann and A. Wool, "Automatic construction of statechart-based anomaly detection models for multi-threaded scada via spectral analysis," in *Proceedings of the 2nd ACM Workshop on Cyber-Physical Systems Security and Privacy*, 2016, pp. 1–12.
37. S. Shukla, S. Thakur, and J. G. Breslin, "Anomaly detection in smart grid network using fc-based blockchain model and linear svm," in *International Conference on Machine Learning, Optimization, and Data Science*. Springer, 2021, pp. 157–171.
38. F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru, and A. Dairi, "Unsupervised deep learning-based process monitoring methods," in *Statistical Process Monitoring Using Advanced Data-Driven and Deep Learning Approaches*. Elsevier, 2021, pp. 193–223.
39. A. Dairi, F. Harrou, Y. Sun, and M. Senouci, "Obstacle detection for intelligent transportation systems using deep stacked autoencoder and k-nearest neighbor scheme," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5122–5132, 2018.
40. A. Dairi, F. Harrou, M. Senouci, and Y. Sun, "Unsupervised obstacle detection in driving environments using deep-learning-based stereovision," *Robotics and Autonomous Systems*, vol. 100, pp. 287–301, 2018.
41. D. Charte, F. Charte, S. García, M. J. del Jesus, and F. Herrera, "A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines," *Information Fusion*, vol. 44, pp. 78–96, 2018.
42. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
43. F. Harrou, F. Kadri, and Y. Sun, "Forecasting of photovoltaic solar power production using lstm approach," *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems*, p. 3, 2020.
44. A. Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting covid-19 time-series data: A comparative study," *Chaos, Solitons & Fractals*, vol. 140, p. 110121, 2020.
45. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
46. L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5046–5063, 2018.
47. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.
48. F. Kadri, A. Dairi, F. Harrou, and Y. Sun, "Towards accurate prediction of patient length of stay at emergency department: a gan-driven deep learning framework," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–15, 2022.
49. R. R. R. Barbosa, R. Sadre, and A. Pras, "Towards periodicity based anomaly detection in scada networks," in *Proceedings of 2012 IEEE 17th International Conference on Emerging Technologies & Factory Automation (ETFA 2012)*. IEEE, 2012, pp. 1–4.
50. H. J. Shin, D.-H. Eom, and S.-S. Kim, "One-class support vector machines—an application in machine fault detection and classification," *Computers & Industrial Engineering*, vol. 48, no. 2, pp. 395–408, 2005.
51. F. Harrou, N. Zerrouki, A. Dairi, Y. Sun, and A. Houacine, "Automatic human fall detection using multiple tri-axial accelerometers," in *2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2021, pp. 74–78.

52. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
53. F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.
54. Liu, Fei Tony and Ting, Kai Ming and Zhou, Zhi-Hua, "Isolation-based anomaly detection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 1, pp. 1–39, 2012.
55. L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
56. P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
57. M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.
58. C.-Y. Lin and S. Nadjm-Tehrani, "A comparative analysis of emulated and real iec-104 spontaneous traffic in power system networks," in *International Workshop on Cyber-Physical Security for Critical Infrastructures Protection*. Springer, 2020, pp. 207–223.
59. G. Clarke, D. Reynders, and E. Wright, *Practical modern SCADA protocols: DNP3, 60870.5 and related systems*. Newnes, 2004.
60. P. Matoušek, O. Ryšavý, and P. Grofčík, "Ics dataset for smart grid anomaly detection," 2022. [Online]. Available: https://dx.doi.org/10.21227/1trw-n685
61. P. Matoušek, V. Havlena, and L. Holík, "Efficient modelling of ics communication for anomaly detection using probabilistic automata," in *2021 IFIP/IEEE International Symposium on Integrated Network Management (IM)*. IEEE, 2021, pp. 81–89.
62. A. Dairi, F. Harrou, Y. Sun, and S. Khadraoui, "Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach," *Applied Sciences*, vol. 10, no. 23, p. 8400, 2020.
63. F. Harrou, Y. Sun, A. S. Hering, M. Madakyaru *et al.*, *Statistical process monitoring using advanced data-driven and deep learning approaches: theory and practical applications*. Elsevier, 2020.