On The Efficacy of Physics-Informed Context-Based Anomaly Detection for Power Systems

Muhammad Nouman Nafees, Neetesh Saxena, and Pete Burnap,
School of Computer Science & Informatics, Cardiff University, Cardiff, United Kingdom
{nafeesm, saxenan4, burnapp}@cardiff.ac.uk

***Abstract*—The Automatic Generation Control (AGC), a fundamental frequency control system, is vulnerable to cyber-physical attacks. Coordinated false data injection attack, aiming to generate fake transient measurements, typically precedes unwarranted actions, inducing frequency excursion, leading to electromechanical swings between generators, blackouts, and costly equipment damage. Unlike other works that focus on point anomaly detection, this work focuses on contextual detection of stealthy cyber-attacks against AGC by utilizing prior information, which is essential for power system operation and situational awareness. More specifically, we depart from the traditional deep learning anomaly detection that is thoroughly driven by black-box detection; instead, we envision an approach based on physics-informed hybrid deep learning detection 'CLDPhy,' which utilizes the combination of prior knowledge of physics and system metrics. Our method, to the extent of our knowledge, is the first context-based anomaly detection for stealthy cyber-physical attacks against the AGC system. We evaluate our approach on an industrial high-class PowerWorld simulated dataset – based on the IEEE 37-bus model. Our experiments observe a 36.4% improvement in accuracy for coordinated attack detection with contextual information, and our approach clearly demonstrates the superiority in comparison with other baselines.**

***Index Terms*—Automatic generation control, coordinated attack, deep learning, context-based anomaly detection**

I. Introduction

Modern power grids have evolved into a more complex cyber-physical system, incorporating integrated communication networks, advanced monitoring and control systems, along with several intelligent hardware devices. With the unique complexity and heterogeneity in the power grid networks comes added vulnerability to emerging threats such as cyber-attacks. For example, nation-state actors mounted one major known attack on the Ukrainian power grid on December 23rd, 2015. Attackers illegally infiltrated Supervisory Control and Data Acquisition (SCADA) systems and computers, culminating in a blackout with catastrophic consequences: a power outage that left 225,000 customers without energy for two to six hours [1]. Consequently, security has become a critical concern, necessitating the development of holistic detection techniques to counter the threats encountered by modern power grid networks.

The vulnerability of the power grid to cyber-attacks is reflected by the demonstration projects, as well as real-world attacks [2], [3]. A significant challenge for security operators is effectively detecting anomalous events: However, not all anomalous events are malicious, culminating in an overwhelming number of false alarms for security operators to investigate. Although security tools such as Security Information and Event Management (SIEM) are effective for alert correlation in Information and Communications Technology (ICT), much of what is known about contextual detection in power grids is still anecdotal.

More specifically, cyber-attacks such as False Data Injection (FDI) attacks against SCADA and Automatic Generation Control System (AGC) have emerged as an important concern [4]. Even worse, besides being afflicted with random load disturbances, adversaries can mount control-based cyber-attack by targeting the Area Control Error (ACE) values sent from the AGC algorithm to the designated generators. On the other hand, advanced adversaries may also manipulate values in the control process loops that collate continuous data via suitable sensors. For example, adversaries can imitate the signatures of a natural load disturbance in the power grid: Worse still, the attack can prevent operators from successfully determining the cause of an anomaly. Consequently, the attack can closely follow the behavior of the physical system. The attackers can make the attack appear to be a plausible physical system behavior until inducing unwarranted actions and causing a severe power outage in the worst scenario.

Malicious anomalies in the SCADA and AGC can be classified as point and contextual anomalies. A point anomaly is defined as an event that differs from its Spatio-temporal neighboring events. In contrast, a contextual anomaly is defined as the events that globally interact in a specific context to cause the unusual manifestation of impacts on the power grid, even if any individual isolated event can be normal. Specifically, context-aware detection can reduce false-positive alerts; for example, they raise suspicion when local and contextual events occur together. Reducing false alarms is crucial; flagging a security event is equally critical when it is not a threat.

A. Related Work

Most research on anomaly detection has focused on detecting point anomalies [5], [6]. Contextual anomalies in an environment may be considered normal or malicious if they arise suspicious under certain conditions in a given context. In context of the AGC, there are two main dual approaches that are mainly used to detect FDI attacks: Model-based and data-driven methods [7]. Model-based methods suffer scalability issues and require extensive construction of power system knowledge-based modeling. Authors in [8] used kernel density estimation to detect FDI attacks in AGC. In [5], load forecasting-based algorithms were utilized for attack detection. However, the detection using the aforementioned method requires a precise forecast of the load profile of the smart grid.

The data-driven methods require historical data, and a training procedure [7]; datasets can be generated through simulations, and there are available historical datasets, which is the motivation to use this approach in this work. Most studies conduct anomaly detection using deep learning to detect FDI attacks in the AGC (see, e.g., [7], [9], [10]). For example, the authors in [10] proposed a neural network-based Luneberger observer to detect attacks in the AGC. Similarly, regression-based predictive models for FDI attacks are proposed in [11], [12]. In this context, authors in [12] developed regression-based signal prediction using long short-term memory (LSTM) networks to detect FDI attacks in the AGC system. One major bottleneck for using deep learning-based techniques is them being 'black-box,' particularly when the model is not equipped with prior knowledge-based contexts.

B. Contributions

Motivated by the aforementioned problems, such as lack of context and prior information in the anomaly detection systems, we propose an approach that involves physics-informed hybrid context-based anomaly detection for the power system by utilizing prior information in the deep learning model. We call it "hybrid" anomaly detection because it combines prior information with multiple neural networks. More specifically, we need to answer the primary research question of this work: *How can deep anomaly detection be utilized to discern point and contextual anomalies in the context of complex cyber-physical attacks?*

In this sense, this paper aims to introduce a context-based anomaly detection framework that exploits prior information in power systems to overcome the bottleneck of black-box detection in such approaches, enabling the system to detect context-based malicious anomalies, even if any individual isolated event appears to be normal. We use the CNN-LSTM-DNN architecture, which we call CLDPhy, to detect anomalies conditioned on a specific AGC context. CLDPhy differentiates from the previous works in how it utilizes the relevant system and physics-informed metrics, inter-area indices of AGC and additional contextual information received from the control center, to boost anomaly detection for stealthy events and enforces properties about how the algorithm identifies anomalous measurements conditioned on the context. As the proposed approach utilizes data already available in most AGC systems, its integration with other detection systems is straightforward. This is crucial, as the detection system would enable more informed response. To evaluate our approach, we use the synthetic datasets from the high-class PowerWorld simulator based on the IEEE 37-bus model. Our results show a 36.4% improvement in accuracy for coordinated attack detection with contextual information. Moreover, beyond simply improving the detection accuracy, our approach produces fewer false positives and higher precision, recall and F1-scores than the other models.

II. System and Attack Models

This work's main idea is to detect stealthy and coordinated attacks that incorporate multiple attack vectors to cause dire consequences to the critical infrastructure such as the power grid. In this section, we discuss the problem statement and detail the system and attack model for this paper.

A. Problem Statement

Coordinated process control loop attacks are the congregation of multiple cyber-physical attack models in which an adversary is assumed to have advanced knowledge of the system's control processes, such as the secondary frequency control mechanism in the power grid [13]. The adversary starts injecting false data into the sensor's values in a coordinated way to keep the attack stealthy: Ensuring the detection statistic remains below the known thresholds. In his journey, from one control loop to another, the adversary continually injects malicious measurements into the critical signals, ensuring not exceeding the threshold values for the various processes.

B. System Model

In this work, we focus on the AGC system of the power grid. The AGC is a wide-area frequency control application that maintains the system frequency at a nominal value (e.g., 60Hz) and keeps the power interchange between Balancing Authority (BA) areas at the scheduled values. The control of the AGC mechanism relies on the closed-loop feedback control system, consisting of sensors, actuators, and controllers (e.g., PLCs). The sensors measure the tie-line power flow between BA areas, bus voltage, and system frequency, typically sent to a control center via an industrial control system known as a SCADA system. The AGC controller computes the control signal known as the (Area Control Error) ACE using these measurements after receiving them over a communication network. For the i^{th} area, $ACE_i = a_i \times P_{E_i} + b_i \times f_i$, where P_{E_i} and f_i are the i^{th} area's power export and frequency deviation of the grid, whereas a_i and b_i are the constants. The ACE values are transmitted to the generators to adjust the primary control loop set-points, and the process is repeated every 2-4 seconds, thus completing the closed loop.

C. Attack Model

Without loss of generality, we consider multiple attack models in the AGC system, introducing various types of FDI attacks in different areas of the AGC system. In this direction, such attacks are not considered once-for-all actions but an iterated process; several iterations of an attack are required to mount an effective attack with an adverse impact [13]. The actions can adversely impact the performance of the AGC system, which can cause generation imbalance and destabilize systems' frequency. From the defender's perspective, such actions can be caused for various reasons unrelated to a cyber-attack; therefore, the absence of context to the anomaly detection in the power system can lead to bad control decisions with dire consequences.

Attack Implementations. This work considers multiple attack scenarios: point anomaly-based basic attacks and contextual anomaly-based stealthily coordinated attacks. For point anomaly-based basic attacks, the adversary mounts an attack on a single measurement signal without any coordination, which can be modeled as:

$$\Delta F_i = a_f(\Delta f_i + S_f), \tag{1}$$

$$\Delta P_t = a_p(\Delta p_t + S_t), \tag{2}$$

where ΔF_i and ΔP_t are the deviations caused by the attack on frequency and tie-line signals, respectively. To this end, a_f and a_p represent the attack factors for frequency and tie-line, respectively. S_f and S_t are the scaling factors with respect to AGC set-points; the increase in values outside the acceptable range can exceed the system's threshold.

For contextual anomaly-based coordinated attacks, the adversary mounts multi-stage attacks involving iterative cycles with coordination, which can be modeled as:

$$\Delta F_i = a_f(t_{min}) \times (\Delta f_i + S_f(t_{min})), \tag{3}$$

$$\Delta P_t = a_p(t_{min}) \times (\Delta p_t + S_t(t_{min})). \tag{4}$$

The adversary mounts a multi-stage attack starting with 20 cycles of a scale attack on tie-line for minimum time in seconds. Simultaneously, the adversary mounts a ramp attack on tie-line 2-3, whereas a random attack is followed on the frequency in area 3. The random attack on the frequency of Area 3 is coordinated with the scale and ramp; it is worth noting that the random attack is used to compensate for the deviation of the sudden frequency change so that the stealthiness of the attack can be maintained. The adversary must ensure that frequency signal, ACE value, and their corresponding rate of change must be within the acceptable range.

III. Context-based Anomaly Detection in Inter-Process Control Loops

In power systems, the effects of the attack on a single area in one control loop must have evident side-effects and connectivity concerning state variables on other areas in control loops of the AGC, and our approach can detect the attack there by utilizing prior information. For example, loss of generation in Area 1 can cause the system frequency and load in other areas to decrease. During the attack, such side-effects on another control loop are inevitable because the adversary can only manipulate the variables for the processes he has already compromised.

A. Workflow of Detection Scheme

The proposed model uses sensor measurements including voltage, frequency, tie-line, and power flow from the AGC time series from PowerWorld. Additional contextual information in conjunction with critical processes and component information is also incorporated for the demonstration of our approach. The model is mainly based on the Convolution Neural Network (CNN), Long Short-Term Memory (LSTM), and Deep Neural Network (DNN) model. Our choice of combining these layers is motivated by [14], which indicates that LSTM performance can be significantly improved by providing better features to the LSTM, which the CNN layers provide, as well as improving output predictions, which the DNN layers provide. The main idea of the model is to feed input features of the power system data, surrounded by temporal context, into a few CNN layers to extract contextual features in conjunction with reducing variations. An attention network with prior knowledge is used to preserve contextual information and assign weights to the power system attributes according to its importance and relevance. The output of the CNN layer is then fed into a few LSTM layers to reduce temporal variations, as overviewed in Algorithm 1. Then, the last LSTM layers' output is fed into a fully connected DNN layer, which utilizes joint learning from labeled data and prior information to predict contextual anomalies in the power system.

Algorithm 1 Initial Process Data Learning Algorithm

```
    Input: Time series data $X = (x_1, x_2, x_3, x_4..., x_t)$;
           Prior information, $P_x$;
           Sliding window lengths, $L_1, L_2$
          Input/output model for CNN-LSTM-DNN;
    Output: fragmented labels in $D$;
            Optimized parameters;
Require: additional information by sliding window $D_y = (d_1, d_2, .., d_x - X + 1)$;
Ensure: Timestamp for data sequence and inter-area correlation;
  Initialize $X_n$,$D_l$,$D_y$;
  for each $Y_n/d_i \in D_y$ do
  Sequence generation $S_i$;
  Spatial feature extraction;
  Temporal feature extraction;
  extract contextual information from sub-network;
  Feed to DNN;
  end for
```

1) Control Invariants:

The proposed anomaly detection for coordinated attacks aims to ensure that each participating area's information is fragmented in the feature space, where specific data features of the measurements are constantly tracked. To identify the strong control invariants for the accuracy of the results, we first employ pattern mining on the dataset in which we get antecedent followed by consequent as the output such that specific antecedent implies relevant consequent.

2) CNN: We utilize the CNN attention block unit to focus on the key features in the time-series data. In so doing, we ensure the unit focuses on the important features in the data and ignores the irrelevant information. The CNN module is formed by multiple layers where each layer has a convolution layer, a non-linear, and a normalization layer with a rectified linear activation function. Furthermore, these layers aggregate samples by employing pooling layers that gradually extract key features via the stacking of convolutional layers.

The employed attention block in our work expands the receptive field of the input, which makes the model capable of attaining contextual information in conjunction with minimizing the interference of irrelevant features to the model. Doing so allows the model to be able to discriminate between the critical information and unimportant information of the time-series data of the AGC.

Prior Information. Separate encoders for prior information and data are employed, allowing it to increase/decrease the strength of additional information and control invariants without retraining. In so doing, two encoders are utilized for the prior information based on control invariants and labeled data measurements to predict anomalies conditioned on the context.

Specific process values are parameterized by variables: *potential violation strength* and *iterative frequency*. The variables have associated scores, which are inspired by the physics-informed process-based metrics. For example, the frequency must not exceed 1Hz during a 15-second window, and the ACE signal must not exceed ±0.05 p.u in potential violation strength variables. Such metrics are indicative of malicious point anomalies; however, these anomalies complement the contextual information. As an example, control center usually receive weather data from weather stations to process load forecasting in accordance with some side-channel metrics. We correlate such metrics and we use the historical values of ACE and tie-lines and power flow measurements as an input for contextual analysis. The integration of prior information is overviewed in Algorithm 2.

3) RNN-LSTM: Neural networks are usually not effective on time series data on account of vanishing gradient problems. A Recurrent Neural Network (RNN) was introduced to overcome such issues, ensuring that the neural network learns the patterns over time. The RNN is capable of predicting sequential data like actions based on previous events. However, using more network layers by RNN makes it challenging to keep track of parameters from the previous layers. Therefore, a variant of RNN, called LSTM, is used to accurately predict sensing time-series data to detect malicious anomalies. LSTM consists of a chain structure with multiple neural network modules, with different gates such as the input gate, output gate, and forget gate; these gates are responsible for selecting or rejecting the information passing through the network.

4) DNN: A deep neural network with multiple layers looks similar to the traditional multi-layer perceptron. They are made up of a layered network structure with a specific number of neurons in each layer. The output layer's node numbers and activation functions are tailored to the classification issue. In addition to the input and output layers, it has hidden layers that can extract complex information. The DNN used in this study contains two dense layers with the rectified linear unit activation function and utilizes joint learning from labeled data and prior information.

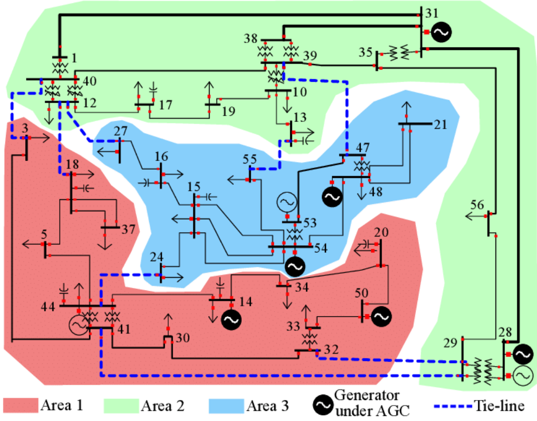

Fig. 1. A three-area 37-bus power grid

Algorithm 2 Integration of Prior Information and Metrics Process

```
Input: training data $D = (x_i, y_i)$;
       Power metrics $M = (m_1, m_2, m_3, ..., m_t)$;
       Threshold, load-profile, spatial-temporal state;
       Correlation between the plurity of AGC attributes;
Output: Adaptive capacity, Spatial-temporal states;
        Optimized parameters;
Require: information encoder $E_p$, trained data encoder $E_d$,
    decision block $B_d$;
Ensure: The label of each fragment correlate with the computed metrics;
Initialize $E_p$,$E_d$,$B_d$;
while not converged do
    Get batch $D$ from $Feeder_x$;
    Get corresponding $M$ for $D$;
    Get features in accordance with sliding window $Wx/E_p/E_d$;
    Update from gradients;
    if any threshold then
        update $Y_s, Q_s$ for $t_x = 0$ to $t_y$;
    end if
end while
```

IV. Experiments and Results

We evaluate our detection approach using simulated datasets for different areas in the AGC system. This section first explores the environment and model settings of our proposed approach. Then, we evaluate our approach using different metrics such as accuracy, precision, recall, and F1-score for point and contextual anomalies. Finally, we compare our results with other baselines.

A. Case Study

We conduct PowerWorld simulations based on the three-area IEEE 37-bus model, an industry-class simulator. Fig. 1 [4] illustrates a three-area grid with 37 buses, where the tie-lines are represented by dotted lines. For reference, we use historical load profiles of NY-ISO [15] to modify the dataset. We consider it an essential step to verify the reliability of the dataset. The states are composed of the voltage, frequency, power flow of the individual buses, tie-line measurements, and the ACE values for various areas. Furthermore, each generator is equipped with 4 second AGC cycle length.

We collect the normal data, and thereafter, we implement false data attacks as reflected in the literature as well as presented in Equations 1-4. Realistic ACE patterns were inserted into this synthetic AGC system. We manually tune the measurements to simulate load fluctuations in multiple areas of the AGC system. The added measurements do not violate any predefined rules for the standard power system scenarios; however, the modification can be anomalous under given conditions based on prior information. To mimic a ramp attack, we carefully inject the attack sequence with respect to ramp up/down generators and deviate frequency and ACE measurements in a coordinated way. The exploitation of vulnerabilities and mounting actual attacks is not the scope of this work. Besides, we realize that combining synthetic and real data can inevitably introduce bias; therefore, we analyze the distribution manually to reduce the bias as much as possible.

B. Context-based Anomaly Detection

We implement our approach using Tensorflow 2.0 with Anaconda 3.0 for programming in Python. We add feeding features into a CNN and perform temporal modeling with an LSTM. We then feed this output into two connected layers of DNN. We also analyze the effect of using different combinations of deep learning. We investigate the impact of adding LSTM before the fully connected layers of DNN. We then compare our approach with several other approaches such as LSTM, CNN-LSTM, and LSTM-DNN and methods that only focus on point anomalies. Sliding windows are used to split the data traces. Power measurements, including sensors and actuation values in conjunction with prior information, are given as input features in our approach. The correlation of information with the attributes ensures the improvement in the model performance in the employed algorithm context.

We trained our model in two settings: with and without prior information context. The "without prior information" setting contains no additional physics-informed metrics and contextual information. The "with prior information" setting is more representative of the information about the control process of AGC, which includes the physics-based system metrics and other attributes relevant to the power system. Furthermore, we evaluate our approach on two sets of attack scenarios: basic FDI attacks and coordinated attacks. The "point anomaly" refers to an anomalous event based on threshold violations. The "contextual anomaly" is a stealthy attack event that looks normal without context. True positives (TP), true negatives (TN), false negatives (FN), and false positives (FP) are used for performance results. These four results are used to compute the metrics for evaluation of our results:

TABLE I
RESULTS AND COMPARISON

Basic FDI Attack with Point Anomaly			
Method	Accuracy	Precision	Recall
LSTM	81.2%	74.4%	72.4%
CNN-LSTM	84.2%	78.1%	76.2%
LSTM-DNN	59.6%	56.2%	51.2%
CLDPhy (contextual with prior)	**96.4%**	**95.3%**	**94.5%**
CLDPhy (non contextual without prior)	**85.2%**	**84.3%**	**80.5%**
Coordinated stealthy FDI Attack with Contextual Anomaly			
LSTM	40.2%	37.5%	34.2%
CNN-LSTM	55.2%	54.5%	53.6%
LSTM-DNN	37.3%	32.4%	28.2%
CLDPhy (contextual with prior)	**93.2%**	**92.5%**	**91.6%**
CLDPhy (non contextual without prior)	**56.8%**	**54.3%**	**52.8%**

$$Accuracy = TP + TN, \tag{5}$$

$$Precision = \frac{TP}{TP + FP}, \tag{6}$$

$$Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall}. \tag{8}$$

We illustrate the results and comparison of our approach in Table 1 and Fig. 2. Our detection approach is superior compared to other models for both point and contextual anomalies. For the basic FDI attacks with point anomalies, CLDPhy with prior information performs significantly better as expected, since the prior information compliments the detection algorithm with more context and accuracy. The models without CNN layers have less accuracy, precision, and recall, which justifies the advantage of having a CNN layer that provides better features for temporal modeling. It is worth noting that increasing window size improved the accuracy and recall results for all the methods; CLDphy clearly outperforms other methods.

CLDPhy performs significantly better compared to all the algorithms against coordinated stealthy attacks with contextual anomalies. Systematically removing prior information from our approach for coordinated attacks observed drastic performance degradation; for example, accuracy decreases from 93.2% to 56.8%, although the performance change in basic FDI detection is not significant (from 96.4% to 85.2%). These results can be attributed to the nature of anomalies and detection approach. For example, the coordinated stealthy attack does not reflect any apparent malicious anomalies; therefore, providing context with additional information boost

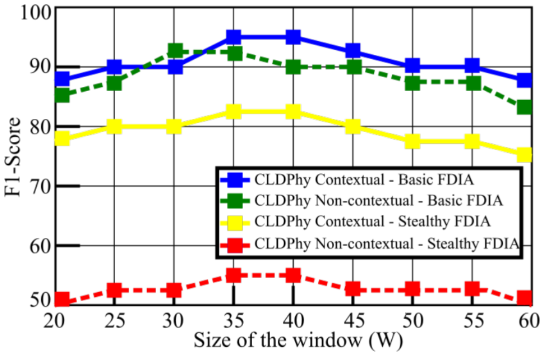

Fig. 2. Comparison results of F1-score under different sliding window lengths

the performance of the detection algorithm. These results support that the prior information-based contextual attributes are vital for the approach to making a conditional inference. More specifically, our approach can achieve high F1-scores; better F1 scores are essential when the FN and FP are crucial. Moreover, F1 is more significant in contextual detection to ensure that specific stealthy attacks are prevented that can cause more damaging effects on the power systems. Fig. 2 shows the F1-score with different window sizes. The CLDphy with contextual information has better F1-scores comparatively. However, performance degrades when the window length is larger than 30 due to the time duration of attacks for basic FDI attacks and point anomalies. In particular, F1-scores for CLDPhy without prior information against coordinated stealthy attacks suffer more degradation. In contrast, CLDPhy with contextual information performs significantly better, and an explanation for this result can be attributed to the contextual anomalies utilization by the algorithm throughout the sliding window lengths. We note that reducing false positives and recall performance can be improved further by optimizing some factors, such as increasing the attack threshold and incorporating attack scenario-specific metrics into the model.

C. Limitations

We realize the limitations of generalizing the results of this paper to other complex cyber-physical system datasets, particularly those with more dynamic power system attributes. Moreover, there might be some practical limitations in terms of incorporating a wide range of physics-informed attributes; the input to the CNN model is manually designed to capture correlations of sensor readings. However, the main idea of our approach is to reduce the false-positive ratio by detecting anomalies based on specific contexts.

V. Conclusion and Future Work

We proposed a context-based anomaly detection approach based on deep learning for the power system. Our strategy to reduce the number of false positives and increase the accuracy, precision, recall and F1-scores involved utilizing hybrid classifiers; a feature extractor and temporal matcher with a given prior information context. Unlike similar works, our approach utilized additional physics-informed metrics from multiple areas of the AGC to detect context-wise anomalies based on prior information, specifically tailored for coordinated attacks. We observe that our proposed approach clearly demonstrates superiority compared to other baselines, and reflects the efficacy of contextual prior-information; for example, CLDphy with contextual prior-information achieved accuracy of 96.4% and 93.2% compared to the accuracy of 85.2% and 56.8% without contextual prior-information for basic FDI and coordinated attacks, respectively.

The proposed approach can be extended to more diverse attack scenarios by incorporating a wide range of prior information based on physics, communication, and control processes. Moreover, individual prior information attributes can be analyzed in terms of the efficacy of the anomaly detection performance. Other aspects can also be considered to improve the efficacy of this work, such as combining sensor measurements and network packets in a hybrid neural network architecture.

References

[1] D. E. Whitehead, K. Owens, D. Gammel, and J. Smith, "Ukraine cyber-induced power outage: Analysis and practical mitigation strategies," in *70th Annual Conference for Protective Relay Engineers*. IEEE, 2017.

[2] M. Zeller, "Myth or reality—does the aurora vulnerability pose a risk to my generator?" in *2011 64th Annual Conference for Protective Relay Engineers*. IEEE, 2011, pp. 130–136.

[3] S. Soltan, P. Mittal, and H. V. Poor, "{BlackIoT}:{IoT} botnet of high wattage devices can disrupt the power grid," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 15–32.

[4] R. Tan, H. H. Nguyen, E. Y. Foo, X. Dong, D. K. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Optimal false data injection attack against automatic generation control in power grids," in *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems*. IEEE, 2016.

[5] S. D. Roy and S. Debbarma, "Detection and mitigation of cyber-attacks on agc systems of low inertia power grid," *IEEE Systems Journal*, vol. 14, no. 2, pp. 2023–2031, 2019.

[6] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. Yao, "Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, pp. 1–36, 2021.

[7] O. Boyaci, M. R. Narimani, K. R. Davis, M. Ismail, T. J. Overbye, and E. Serpedin, "Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 807–819, 2021.

[8] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[9] F. Zhang and Q. Li, "Deep learning-based data forgery detection in automatic generation control," in *2017 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2017, pp. 400–404.

[10] A. Abbaspour, A. Sargolzaei, P. Forouzannezhad, K. K. Yen, and A. I. Sarwat, "Resilient control design for load frequency control system under false data injection attacks," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7951–7962, 2019.

[11] M. N. Nafees, N. Saxena, and P. Burnap, "Optimized predictive control for agc cyber resiliency," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 2450–2452.

[12] C. Chen, Y. Chen, J. Zhao, K. Zhang, M. Ni, and B. Ren, "Data-driven resilient automatic generation control against false data injection attacks," *IEEE Transactions on Industrial Informatics*, 2021.

[13] X. He, X. Liu, and P. Li, "Coordinated false data injection attacks in agc system and its countermeasure," *IEEE Access*, 2020.

[14] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks," *arXiv preprint arXiv:1312.6026*, 2013.

[15] N. ISO, "Load data." [Online]. Available: https://www.nyiso.com/load-data