Physics-Informed Augmentation for Contextual Anomaly Detection in Smart Grid

Muhammad Nouman Nafees
Cardiff University
Cardiff, United Kingdom
nafeesm@cardiff.ac.uk

Neetesh Saxena
Cardiff University
Cardiff, United Kingdom
saxenan4@cardiff.ac.uk

Pete Burnap
Cardiff University
Cardiff, United Kingdom
burnapp@cardiff.ac.uk

ABSTRACT

Smart Grid (SG) networks, as a part of critical national infrastructure, are vulnerable to sophisticated cyber-physical attacks. Specifically, a coordinated false data injection attack aiming to generate fake transient measurements in the SG's Automatic Generation Control (AGC), can cause unwarranted actions and blackouts in the worst scenario. Unlike other works that overlook contextual correlations, this work utilizes prior information and a temporal model to detect cyber-attacks. Specifically, we depart from the traditional deep learning anomaly detection, driven by black-box detection; instead, we envision an approach based on physics-informed hybrid deep learning detection. Our approach utilizes the combination of process control-based variational autoencoder, prior knowledge of physics, and long short-term memory for a false data injection attack. To the best of our knowledge, our method is the first contextual-based anomaly detection that incorporates process control prior information against complex attacks in the smart grid. The proposed approach is evaluated on the modified high-class PowerWorld simulated dataset based on the IEEE 37-bus model. Our experiments observe the lowest reconstruction error and offer 96.9% accuracy, demonstrating superiority over other baselines.

CCS CONCEPTS

• **Security and privacy** → **Intrusion detection systems**;

KEYWORDS

automatic generation control; contextual detection; deep learning

ACM Reference Format:
Muhammad Nouman Nafees, Neetesh Saxena, and Pete Burnap. 2022. Physics-Informed Augmentation for Contextual Anomaly Detection in Smart Grid. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), November 7–11, 2022, Los Angeles, U.S.A.*. ACM,NewYork,NY,USA,3pages.https://doi.org/xx.xxxx/xxxx.xxxx

1 INTRODUCTION

Smart grid (SG) networks orchestrate advanced communication, sensing, monitoring, and control technologies. With these integrations come added vulnerability to emerging threats knowns as cyber-attacks. More specifically, coordinated False Data Injection (FDI) attacks have emerged as an important concern [4]. Besides random load disturbances, the adversary can simultaneously falsify power system measurements that collate data via suitable sensors to mislead control operations maliciously. Compared with traditional FDI attacks, this simultaneous attack can closely follow the behavior of the physical system to vitiate the detection mechanism. Worse still, the attackers can prevent operators from successfully determining the cause of an anomaly, potentially inducing unwarranted actions to cause a power outage in the worst scenario. Consequently, cyber-security has become a significant concern, necessitating the development of holistic detection techniques to counter the threats encountered by modern power grid networks.

The prevalent detection approaches for FDI attacks are broadly classified into model-based and data-driven approaches. Model-based methods require a deep understanding of the underlying power-system modeling and suffer scalability issues [2]. The data-driven methods require historical data and a training procedure; datasets can be generated through simulations, and there are available historical datasets, which is the motivation to use this approach in this work. However, deriving complex physical dynamics based on only data is still beyond the scope of Machine Learning (ML), and Deep Learning (DL) approaches.

While there have been efforts to incorporate temporal correlations in attack detection [1, 3], none of these efforts consider the physics-informed and process control-based prior knowledge between input and output variables and control invariants: It is particularly crucial for attacks where the adversaries can imitate the signatures of natural load disturbance in the power grid. For example, on a power system dataset with only point anomalies to reflect load disturbance, a deep learning model may learn that such anomaly is associated with a particular load disturbance. While this may be a valid correlation, the crux, rather, is that it does not reflect the true effect between input and output and will lead to incorrect predictions on test data that include more stealthy attack instances with varying impacts on the control processes of the SG. Therefore, much of what is known about contextual detection in power grids is still anecdotal.

Contributions. Motivated by the aforementioned challenges, we propose a hybrid approach coupling data-driven deep learning versatility and physics-informed prior information for power grid networks. More specifically, we need to answer the primary research question of this work: *How can deep anomaly detection be utilized to augment physics-informed attributes to reflect the temporal correlations in the data-driven approach for contextual detection in power systems?*

CCS '22, November 7–11, 2022, Los Angeles, U.S.A.
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-8454-4/21/11...$15.00
https://doi.org/xx.xxxx/xxxx.xxxx

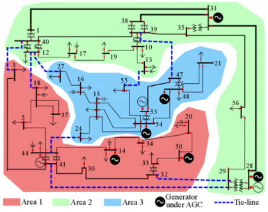

Figure 1: A three-area 37-bus power grid

In this sense, we utilize the prior information based on the observed data and auxiliary elements in the Automatic Generation Control (AGC) of SG to detect coordinated FDI attacks. To this end, prior knowledge can augment the training data containing process control and correlational relationships between certain input variables and invariants. In so doing, we decompose the information into a data-driven term such that the process control-based knowledge only compliments the information that the data-driven approach cannot reflect.

Going beyond, we aim to introduce a deep learning-based context-aware detection framework that augments physics-informed attributes driven by the deep learning method. As such, our approach enforces hybrid properties to ensure interpretability and generalization in conjunction with minimizing the bottleneck of black-box detection in power systems. We use the Process Control Variational Autoencoder (VAE) with a context sub-network and Long-short Term Memory (LSTM), which we call PC-VAE-LSTM, to detect cyber-attacks conditioned on context. To evaluate our approach, we use the synthetic datasets from the high-class PowerWorld simulator based on the IEEE 37-bus model. Beyond improving detection accuracy, our approach produces fewer false positives and higher precision, recall, and F1 scores than the other models.

2 PHYSICS-INFORMED CONTEXTUAL DETECTION

This section presents the AGC and attack model details. Furthermore, we provide the workflow of the detection scheme, seeking to reflect the temporal correlation with physics-informed prior information.

2.1 System and Attack Model

System Model. We consider a standard electric power AGC system, where the system's function is to maintain the system frequency at its nominal value, e.g., 60 Hz in North America, and to regulate the net scheduled value of power flow across different Balancing Authorities (BAs) [4]. The AGC controller computes the Area Control Error (ACE), an integral part of the AGC algorithm, using these measurements after receiving them over a communication network. For the i^{th} area, $ACE_i = a_i \cdot P_{E_i} + b_i \cdot f_i$, 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 sent to the generators to adjust the primary control loop set-points, and the process is repeated every 2-4 seconds, also referred to as the AGC cycle. Figure 1 [6] illustrates a three-area grid with 37 buses, where the tie-lines are represented by dotted lines.

Attack Model. We consider a coordinated attack model in the AGC system, introducing multi-staged stealthy FDI attacks with several coordinated iterations in different areas of the AGC system. The adversary mounts a multi-staged attack starting with 20 cycles of a scale attack on a tie-line for a minimum time of seconds. Simultaneously, a ramp attack on tie-line 2-3 is executed, 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 attack; it is to compensate for the deviation of the sudden frequency change to ensure that the frequency signal, ACE value, and their corresponding rate of change must be within the acceptable range.

2.2 Workflow of Detection Scheme

Network Architecture. Our approach uses CP-VAE-LSTM, a fairly simple network structure, and unlabeled training requirements, with a context subnetwork to explicitly incorporate physics-informed prior information concerning the AGC mechanism. The context subnetwork acts as a variant of process-specific control logic to tune the weight added in the VAE during the output divergence as feedback during model training. To enrich this model with temporal correlations, we use the LSTM encoder-decoder. According to the observations, the ACE data of control areas have specific patterns based on the physical configurations of the AGC system. Any manipulations will disrupt these patterns. Therefore, a Recurrent Neural Network (RNN) variant, called LSTM, is used to predict sensing time-series data accurately and detect malicious anomalies. The framework takes initial input from sensor measurements, including voltage, frequency, tie-line, and power flow from the AGC time series PowerWorld generated dataset. The anomaly predictions are filtered via the contextual subnetwork and the LSTM-based encoders. The detection approach proceeds in two steps.

In this first step, the model uses the reconstruction error to detect anomalies in data. During the inference, it takes sensor and process control measurements, including anomalies, and returns the corresponding latent variable $X = (x_1, x_2, ., x_t)$ as outputs. Using the output, it tries to calculate reconstruction. As such, for anomalies, the reconstruction error is higher than the normal data. Since it uses the reconstruction error as an anomaly score, a high reconstruction performance is necessary to prevent false load deviation predictions concerning what is normal or abnormal.

Prior Information. The second step is crucial to augment prior information in the context subnetwork and convert unsupervised training mode into semi-supervised training mode. The motivation behind this is that we encode some additional information into the model if there is knowledge of certain links. If, for example, we know that one variable is dependent on another, we can compute that variable in a certain way that makes sense concerning the dependent variables. We modify the training approach by introducing the prior information context subnetwork so that the output of the subnetwork and LSTM encoder-decoder is obtained as a single representation 'Z'.

The variables have associated scores inspired by the physics-informed process-based metrics. For instance, 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.

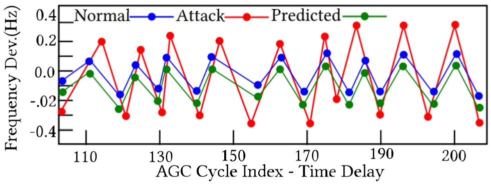

Figure 2: Predicted frequency deviation

We correlate such metrics and use the historical values of ACE, tie-lines, and power flow measurements as an input for contextual analysis. The contextual attributes such as line tripping data, and load change profile in conjunction with date/time are additionally given as vectors to the context sub-network. The correlations of loads can reflect this dependency at different times, which also implies the correlations on states and measurements.

3 EVALUATION

Dataset. We conduct PowerWorld simulations based on the three-area IEEE 37-bus model, an industry-class simulator. For reference and reasonable load profile, we use historical load profiles of NY-ISO [5] to modify the dataset. The states are composed of the individual buses' frequency, voltage, power flow, tie-line measurements, and ACE values. Furthermore, each generator is equipped with 4 second AGC cycle length. We manually tune the measurements in a coordinated way to simulate load fluctuations in multiple areas of the AGC system without violating any predefined rules for the standard power system scenarios; however, the modification can be anomalous under given conditions based on prior information. Exploiting vulnerabilities and mounting actual attacks is not the scope of this work.

Contextual Detection. We add feeding features into a PC-VAE and perform temporal modeling with LSTM encoders. 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 to identify to what extent the model degrades without the additional information. The "with prior information" setting is 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. We then compare our approach with several other baselines, including AE, VAE, and VAE-LSTM. True positives (TP), true Negatives (TN), False Negatives (FN), and False Positives (FP) are used for performance results. These four results are computed for the Accuracy, Precision, Recall, and F1 scores.

Results. Table 1 and Figure 2 show the performance of our approach in terms of given evaluation metrics and load deviation prediction in AGC, respectively. The algorithm successfully reconstructs the frequency deviation based on prior information and closely follows the expected deviation in normal settings. PC-VAE-LSMT is superior compared with their respective baselines. Overall, AE and VAE have relatively poor performance; these algorithms do not exploit the temporal information nor leverage the prior information between the variable and the control invariants. Therefore, it justifies the advantage of utilizing LSTM, which provides better features for temporal correlation. However, removing prior information from our approach observed drastic performance degradation for contextual anomaly detection. We observed slightly better performance for iterative coordinated FDI attacks; the flip side is significantly higher latency in anomaly detection. Unsurprisingly, PC-VAE-LSTM performs significantly better with 96.9% accuracy, 95.3% precision, 93.4% recall, and 94.3% F1 score. Moreover, the reconstruction loss of our approach also outperforms other algorithms. We note that reducing false positives and recall performance can be improved further by optimizing some factors, for example, choosing the better attack threshold and incorporating attack scenario-specific control invariants and better prior information into the model.

Table 1: Results and Comparison

Method	Acc.	Prec.	Rec.	F1	Recon. error
AE	78.3%	76.1%	74.2%	75.1%	0.09
VAE	84.2%	82.9%	79.2%	81.0%	0.05
VAE-LSTM (without prior)	88.4%	86.5%	85.6%	86.0%	0.03
PC-VAE-LSTM (contextual with prior)	**96.9%**	**95.3%**	**93.4%**	**94.3%**	**0.00**

Acc.: Accuracy; Prec.: Precision; Rec.: Recall

4 CONCLUSION AND FUTURE WORK

We proposed a context-based anomaly detection approach based on hybrid deep learning for power systems. Unlike previous work that overlooked underlying prior information and control invariant correlation, this paper presented a new physics-informed augmentation-based approach that exploits temporal and contextual correlation to detect coordinated FDI attacks in power systems. We observe that our proposed approach is superior to other baselines; for example, PC-VAE-LSTM with prior contextual information achieved 96.9% accuracy, 95.3% precision, 93.4% recall, and 94.3% F1 score. Moreover, the reconstruction loss of our approach also outperforms other algorithms. For future work, we plan to build on the current work by utilizing more prior information, including sensor measurement and network packets in a hybrid neural network, to further enhance contextual detection in power systems.

REFERENCES

[1] Abdelrahman Ayad, Mohsen Khalaf, and Ehab El-Saadany. 2018. Detection of false data injection attacks in automatic generation control systems considering system nonlinearities. In *2018 IEEE EPEC conference*. IEEE.

[2] Osman Boyaci, Mohammad Rasoul Narimani, Katherine R Davis, and Ismail. 2021. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. *IEEE Transactions on Smart Grid* (2021).

[3] Chunyu Chen, Yang Chen, Junbo Zhao, Kaifeng Zhang, Ming Ni, and Bixing Ren. 2021. Data-driven resilient automatic generation control against false data injection attacks. *IEEE Transactions on Industrial Informatics* (2021).

[4] Xi He, Xuan Liu, and Peng Li. 2020. Coordinated false data injection attacks in AGC system and its countermeasure. *IEEE Access* 8 (2020), 194640–194651.

[5] NewYork ISO. 2019. LOAD DATA. (2019). https://www.nyiso.com/load-data

[6] Rui Tan, Hoang Hai Nguyen, and Foo. 2016. Optimal false data injection attack against automatic generation control in power grids. In *2016 ACM/IEEE 7th ICCPS*. IEEE.