Data-Driven Probabilistic Anomaly Detection for Electricity Market under Cyber Attacks

Mucun Sun, Lingyu Ren, Nai-yuan Chiang
United Technologies Research Center (UTRC)
East Hartford, CT 06108, USA
Email: mucunsun@gmail.com

***Abstract*—Information and communication technologies have been widely used in smart grid for efficient operation. However, these technologies are vulnerable to malicious cyber attacks, which may lead to severe reliability and economic issues. Recently, a variety of data-driven anomaly detection approaches have been explored to detect potential cyber attacks in smart grids. In this paper, we researched on the electricity market data aiming to identify anomalies from the locational marginal prices (LMPs) and provide a new indicator for potential cyber attacks in power grids. Specifically, a novel data-driven probabilistic anomaly detection framework is proposed for electricity market, which consists of three major components: long short-term memory (LSTM) based deterministic electricity price forecasting, probabilistic electricity price forecasting and anomaly detection. This framework is tested on a model-based electricity market simulator under two types of cyber attacks, i.e., load redistribution attack (LRA) and price responsive attack (PRA). Numerical results on the simulated LMPs show that the proposed framework is capable of detecting data anomalies over these attacks.**

***Index Terms*—cyber security, anomaly detection, machine learning, probabilistic forecasting, locational marginal price (LMP).**

I. Introduction

The U.S. power grid is a complex cyber-physical system incorporating vast volume of distributed devices, which by nature results in a large attack surface. Malicious attackers can compromise the power devices, communication and control facilities or market interfaces, leading to local outages, equipment damages, grid instabilities or individual financial gains. One of the most cited cyber attacks in literatures is the false data injection attack (FDIA). Motivated by financial benefits, using FDIA, the attackers can create false load estimation or transmission congestion limits to mislead the real-time electricity pricing algorithms in producing biased locational marginal prices (LMPs). The attacker then takes advantages of the biased LMP to gain monetary profits using bids and offers via the market interface. Such coordinated attack is hard to detect using traditional cyber defense methods, such as bad data detection in state estimation and intrusion detection systems in SCADA. The reason is twofold: (1) the current FDIA can be designed sophisticated enough to bypass the bad data detection algorithms [1]; (2) the FDIA can be applied to various data interfaces, such as field sensors, data server in control centers or remote communication channels, using various techniques, e.g. man-in-the-middle attack and malware attack. The complexity and uncertainty of FDIA makes it unreliable to use single focused cyber defense mechanism.

Due to the wide adoption of load management technologies, the research on load alternating attacks (LAA) gains popularity. Three types of LAA were discussed in [2] targeting on (1) data centers and computation load, (2) direct load control and (3) indirect load control. The authors also presented a cost-efficient protection strategy by selecting the critical loads to be guarded based on protection costs, network topology and generation reserves. [3] presented a dynamic load alternating attack (D-LAA) that manipulates loads progressively to destabilize or degrade the system frequency regulation. D-LAA can be designed in open-loop or close-loop using frequency sensor data as the feedback signal. In [4], the authors defined the Manipulation of demand via IoT (MadIoT) attacks, which maliciously control the power demand by compromising IoT devices, such as smart thermostats. With large amount of victim devices, this attack could cause (1) frequency instability, (2) cascading failure and (3) operating cost increase. Similarly, in [5], the authors presented coordinated load-changing attacks, called Grid Shock, targeting on digital devices, i.e. computers and their peripherals. Each of these devices only contribute hundreds to thousands of Watts but a botnet can joint these power consumptions to affect the grid operations negatively. It is noexpensive and almost infeasible to defend LAAs by monitoring and protecting each grid edge devices.

A promising cyber defense approach, which is non-intrusive to the operational system and adds additional protection is the physical response based anomaly detection. For cyber-physical systems, evaluating physical performance from sensor data is a common practice, but using these data to detect cyber attacks is under-developed. A few anomaly detection technologies have been presented in the literature with power systems applications. For example, Wang *et al.* [6] presented a power consumption anomaly detection method based on long short-term memory (LSTM) point forecasts and error pattern. In [7], Krishna *et al.* adopted Principal Component Analysis (PCA) and density-based spatial clustering on noise pattern to detect the anomalies which are deviations from the normal electricity consumption behavior. Kim *et al.* [8] presented a framework which utilizes spatial and temporal correlation between multiple solar farms to defend against data integrity attacks and learns the inter-farm/intra-farm correlation between measurements to perform anomaly detection.

However, most of the existing anomaly detection applications for power systems are deterministic and thus insufficient to characterize the uncertainties of cyber attacks. Probabilistic approaches that provide quantitative uncertainty information associated with cyber attacks are therefore expected to better assist power system operations.

To address the aforementioned limitations, in this paper, a data-driven probabilistic anomaly detection methodology is developed to provide reliable defense strategies against various cyber attack scenarios. First, a deep neural network, LSTM, is used to model the temporal dependencies within the LMP profile and correlations with explanatory variables. Then, a parametric probabilistic forecasting model is adopted to convert the LMP point forecasts to probabilistic forecasts, which is used for anomaly detection. Our major contribution is to formulate the anomaly detection problem as a probabilistic forecasting task and implement this approach to the publicly available electricity market data. The proposed probabilistic anomaly detection algorithm utilizes prediction interval to reveal the underlying structures within normal behavior and detect unexpected events.

The rest of the paper is organized as follows. Section II describes the proposed probabilistic anomaly detection method, in consist of a deep-learning based deterministic forecasting model and a parametric probabilistic forecasting model. Section III applies and validates the developed probabilistic anomaly detection method to two types of cyber attack scenarios. Concluding remarks and future work are discussed in Section IV.

II. Methodology

The overall framework of the proposed probabilistic anomaly detection methodology is illustrated in Fig. 1. It consists of three major steps:

1) Step 1 (gray blocks): Feed the historical data into an LSTM based forecasting machine to predict LMP.
2) Step 2 (orange blocks): Convert the point forecasts to prediction intervals (PIs) using a parametric probabilistic forecasting method based on designated predictive distribution shapes and pinball loss optimization.
3) Step 3 (blue block): Detect anomalies based on the threshold confidence and evaluate the performance.

This section delineates the mathematical models of algorithms used in our proposed framework. Step 1 is explained in Subsection A and Step 2&3 are explained in Subsection B.

A. Multi-input Long Short Term Memory

Due to data availability, it is impractical to collect all explanatory variables (e.g., temperature and humidity, etc.) to build an ideal LMP forecasting model. In this paper, we select the energy cost, congestion cost, forecast load, and their corresponding lagged variables to train the LSTM model, since they are published in real-time for most electricity market operators.

LSTM is a special recurrent neural network (RNN) architecture for time series modeling and forecasting, which has the capability of learning and memorizing long-term dependencies within the time-series data. The basic topology of standard RNN is shown in Fig. 2, where X denotes input, Y denotes output. h is the hidden state, W_{hx}, W_{yh}, and W_{hh} are the weight matrix among inputs, outputs, and hidden state itself, respectively. The standard RNN have one hidden layer, which could only trace back to few time steps due to the vanishing gradient effect [9]. To better capture the long-term dependencies, LSTM introduces different gates which could regulate the gradient flow of the network. Following the work of [10], the inner structure of the LSTM unit is illustrated in Fig. 3 and described in Eq.1.

$$\begin{aligned} i_t &= \sigma(x_t W_{ix} + h_{t-1} W_{ih} + c_{t-1} W_{ic} + b_i) \\ f_t &= \sigma(x_t W_{fx} + h_{t-1} W_{fh} + c_{t-1} W_{fc} + b_f) \\ c_t &= c_{t-1} f_t + i_t \cdot tanh(x_t W_{xc} + h_{t-1} W_{ch} + b_c) \\ o_t &= \sigma(x_t W_{ox} + h_{t-1} W_{oh} + b_o) \\ h_t &= o_t \cdot tanh(c_t) \end{aligned} \tag{1}$$

where $i_{(\cdot)}$, $f_{(\cdot)}$, and $o_{(\cdot)}$ are the input gate, forget gate, and output gate, respectively. σ denotes the sigmoid activation function, h_t is the state at t, x_t denotes input, o_t is the cell output, and c_t is the memory state. LSTM updates its hidden state c_t by using the current input x_t and the previous state c_{t-1}. The final state h_t is determined by c_t and o_t. The weights are optimized by minimizing the difference between the LSTM outputs and training samples. In this study, the input vector of the multi-input LSTM can be expressed as:

$$x_t = [y_{t-1}, \Phi_t] \tag{2}$$

where Φ_t denotes the feature vector of the time step t, y denotes the observation at time step t.

B. Probabilistic Anomaly Detection

Once the deterministic LMP forecasts are generated, a multi-distribution database is formulated to model the possible shapes of the LMP predictive distribution. These four distributions, characterized by mean value (μ) and standard deviation (σ), are Gaussian, Gamma, Laplace and non-central-t distributions. The mean value is approximated by the deterministic point forecast and the standard deviation σ is calculated by minimizing the pinball loss of the quantile function at each time step. Based on the optimal pinball loss values, we select the best predictive distribution. The pinball loss value of a certain quantile L_m is expressed as:

$$L_{m,t}(q_{m,t}, y_t) = \begin{cases} (1 - \dfrac{m}{100}) \times (q_{m,t} - y_t), & y_t < q_{m,t} \\ \dfrac{m}{100} \times (y_t - q_{m,t}), & y_t \geq q_{m,t} \end{cases} \tag{3}$$

where y_t represents the tth observation, m represents a quantile percentage from 1 to 99, and q_m represents the predicted quantile. For a given m percentage, the quantile q_m represents the value of a random variable whose cumulative distribution function (CDF) is m percentage. Pinball loss is one of the most popular metrics for evaluating probabilistic forecasts

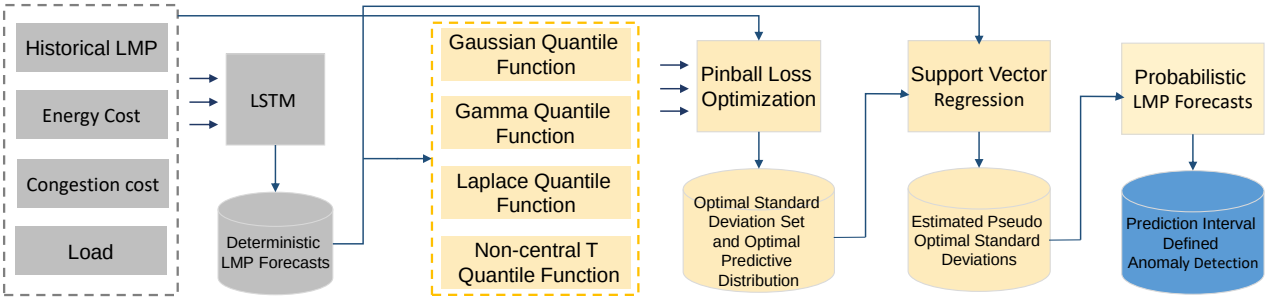

Fig. 1: The Overall framework of the probabilistic anomaly detection model for electricity market data

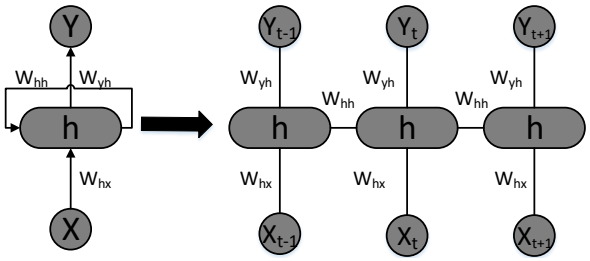

Fig. 2: The structure of RNN

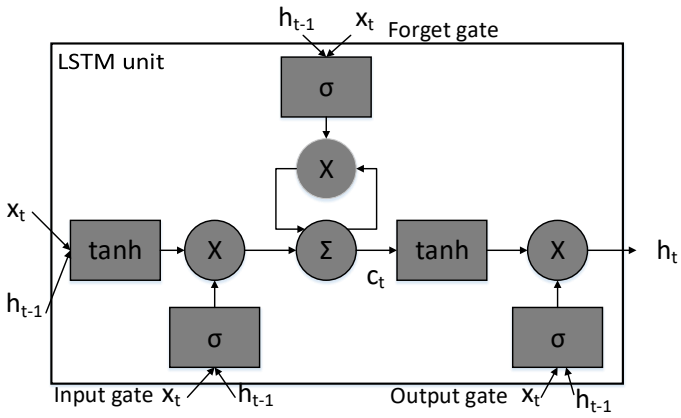

Fig. 3: The inner structure of LSTM unit

[11]. Smaller pinball loss values indicate better probabilistic forecasting.

The process of probabilistic anomaly detection is described as follows:

1) Parameterizing the quantile in terms of μ and σ, where μ assumes to be the point forecast. The mth quantile of the tth point forecast, $q_{m,t}$ is expressed as:

$$q_{m,t} = F^{-1}(\frac{m}{100}, \hat{y}_t, \sigma_t) \tag{4}$$

where, $\hat{y}_t$ and y_t are deterministic forecasts and observations, respectively. $F^{-1}(\cdot)$ is the inverse CDF function. The corresponding pinball loss is expressed as Eq. 3.

2) Calculating the unknown parameter σ at each time step by minimizing the averaged sum of pinball loss through genetic algorithm (GA) [12]:

$$\sigma_t^* = \arg\min_{\sigma_t} \frac{1}{N_m} \sum_{m=1}^{N_m} L_{m,t}(\sigma_t, y_t, \hat{y}_t, m)$$
$$subject\ to \qquad \sigma_l \le \sigma_t \le \sigma_u \tag{5}$$

where σ_t^* is the optimal standard deviation of the tth time step; N_m = 99 is the number of quantiles; σ_l and σ_u are the lower and upper bound of σ, which are set as 0.01 and 80, respectively.

3) An support vector regression (SVR) surrogate model [13] is used to fit the point forecast and σ^* in the training stage, which is used to generate unknown pseudo optimal standard deviations, $\hat{\sigma}^*$, in the forecasting stage.
4) During probabilistic forecasting, both the deterministic forecasts, i.e. μ, generated by LSMT and the estimated pseudo optimal standard deviation $\hat{\sigma}^*$ generated by SVR are used to determine the prediction interval (PI) [14].
5) For each time step, the observation falls into a certain PI, which is used to estimate the likelihood of it being anomaly (outliers). The deterministic prediction decides the best estimate of next step LMP, while the probabilistic prediction quantifies the uncertainty of all possible observations. The larger PI denotes further deviation from its nominal value. In this paper, we assume an anomaly is spotted whenever the observation falls out of the 70% PI. This detection threshold can be further tuned through a sensitivity study which is out of our scope.

III. Case Study

A. Data Description

For data preparation, we first built an electricity market simulator based on Matpower [15] using a combined model of day-ahead economic dispatch (ED) and real-time incremental economic dispatch (IED). We then run the simulator on the IEEE 14 bus system with 11 loads selected from PJM load profiles. The day-ahead hourly load forecast and 5-minute real-time load forecast from Sept. 19th to Oct. 17th 2019 were used

for ED and IED, separately. The simulated 5-minute real-time LMP, energy cost, congestion cost and the 5-min load forecast were used for training and testing. The ratio of the number of training samples to testing samples was 3:1. The LSTM deterministic forecasting model has two hidden layers of 50 and 30 neurons and the weights were optimized with Adam.

B. Cyber Attack Scenario Design

Two kinds of attacks are implemented in the simulator: Load Redistribution Attack (LRA) and Price Responsive Attack (PRA).

LRA was first introduced by Yuan et al [16], as a kind of FDIA attack where only the measurements related to some load bus power injection are attacked. LRA redistributes the load by increasing/decreasing certain loads at some buses while keeping the total load unchanged [17]. Since no attacks happened on the well-protected generation buses and LRA can bypass bad data detection, LRA can be hard to detect in real-time. The damage of LRA is that it can lead to a wrong dispatch result, i.e., fake solution from the economic dispatch problem, which may overload certain transmission lines and raise LMPs. The LRA was added to the simulator before solving the IED problem to redistribute the 5-min load forecast. It tries to increase the load prediction in the targeted bus, while decrease the load prediction in the non-targeted buses. To maximize the gain of the attacks, LRA is only activated in the simulator during the critical hours, i.e., when LMP has a big change in the historical data. In our case, we observed that LMP changes dramatically during 11:00 to 13:00, when LMP increases due to line congestion, and 20:00 to 22:00, when LMP decreases due to the remove of congestion. Therefore, we only add attacks during these two time periods, to extend the time of line congestion. LRA is added by the following procedure:

1) Within the time period 11:00 to 13:00, check if we have already applied attacks in the previous steps. If the number of existing attacks is greater than the maximum allowed attacks, terminate.
2) Check if the next total load prediction is greater than the current total load by 5%. If yes, we reduce the load increasing rate at the non-targeted bus where its corresponding load is increasing, and make the load decreasing rate higher at the non-targeted bus where its load is decreasing. We then apply the adjusted load to the targeted bus to increase its incremental load and accelerate the LMP ramping.

Similar procedure is applied to the time period 20:00 to 22:00, where we aim to slow down the load decreasing rate at the targeted bus, in order to extend the period of line congestion.

PRA is a type of LAA, inspired by the real-time pricing attacks [18], and Manipulation of Demand attack (MAD) [19], which change the load behaviors to damage the power grid. The motivation of our PRA is that the quick growth of smart grid foresees the wide usage of load management technologies, which can change the load behaviors based on the current LMP information. For example, the controller of smart appliance can switch to the full-power mode when the price is low, and keep in energy saving mode when the price is high. Unlike the infrastructures of power grid, which are well-protected, the load controllers are located in the user end with much less secure to defend cyber attacks. The PRA is designed by injecting false price signal to the load controllers so as to inverse the controller logic, to use more power when LMP is high and there is a high opportunity of line congestion in the power grid. By increasing the load demand at such a critical time period, we expect it can possibly change the LMP by introducing more congestion.

Assuming there is a delay in the control of price-responsive demand after LMP changes, e.g., the load change happens 30 minutes after the price change. Our implementation of PRA is summarized as follows:

1) Check if we have already applied attacks in the previous steps. If the number of existing attacks is greater than the maximum allowed attacks in the given time period, terminate. (In our test case, we check if there are 5 continuous attacks happened during the last 5 time steps.)
2) Check if LMP has a big increase, compared to the LMP from last step. (In our test case, we check if it has an increase over 10%.) If yes, adjust the price-responsive load by the following equation

$$PD_{adjusted} = PD_{base} + PD_{pr} * (\lambda_{curr}/\lambda_{pred})^{\beta} \quad (6)$$

where PD_{base} and PD_{pr} are the base load and the price-responsive part of the load, similar to the definition given in [18]. In our test case, we set $PD_{base} = 90\% * PD_{pred}$, i.e., 90% of the load forecast, and the rests are the price-responsive load. The decreasing factor β is set to be -0.8. Note that our formulation is slightly different from the one used in [18], since we use day-ahead LMP prediction λ_{pred} as a base line. If the current LMP is equal to the predicted one, the above equation ends up to $PD_{adjusted} = PD_{pred}$.

C. Deterministic LMP Forecasting Results

Three evaluation metrics are used to assess the deterministic forecasting accuracy, which are the normalized root mean squared error (nRMSE), normalized mean absolute error (nMAE), and mean absolute percentage error (MAPE). For these metrics, a smaller value indicates better forecasting performance. Deterministic LMP forecasting results are summarized in Table I. In this study, the persistence method (PS) is adopted as the baseline since its superiority for shorter forecast horizon [20]. Overall, the accuracies of the LSTM deterministic LMP forecasts are better than those of persistence forecasts under both cases with or without attack. It is mainly because the LMP data is highly temporal correlated, and the LSTM model outperforms in capturing long-term dependencies.

D. Probabilistic Anomaly Detection Results

This section evaluated the performance of the proposed probabilistic anomaly detection method. Laplace distribution is selected as the predictive distribution based on its minimal

TABLE I: 5-min ahead LMP forecasting performance

Model	Metric	Scenario		
		w/o attack	LRA	PRA
LSTM	NMAE(%)	1.07	5.20	5.96
	NRMSE(%)	1.35	6.10	6.83
	MAPE(%)	2.28	7.99	8.43
PS	NMAE(%)	3.80	6.40	6.66
	NRMSE(%)	6.50	7.86	8.42
	MAPE(%)	4.71	9.13	10.27

pinball loss in the training process. Therefore, the LSTM model with Laplace distribution (LSTM-Laplace) is chosen as the final anomaly detection model. The performance skill scores could be calculated based on Table II, where The true positive (TP) denotes the number of detected attacks; false negative (FN), i.e., type II error, denotes the number of missed detection of attacks; False positive (FP), i.e, false alarm or type I error, denotes the number of normal data is treated as attacks; true negative (TN) denotes the number of normal data is correctly identified. N_s is the total number of test samples. Among these indexes, the FP can cause false alarms, which may add redundant work to system operators, while the FN missed by the detection model may bring loss to market end users.

TABLE II: Contingency table of attack detection

	Attack (Yes)	Attack (No)	Total
Detected (Yes)	TP (hit)	FP (miss)	TP+FP
Detected (No)	FN (miss)	TN (hit)	FN+TN
Total	TP+FN	FP+TN	N_s=TP+FP+FN+TN

1) Evaluation Metrics: We calculated the true positive rate (TPR), false positive rate (FPR), and F1 score of the anomaly detection results. The mathematical expressions of the three metrics are expressed as:

$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

$$F\text{-}1 = \frac{2TP}{2TP + FP + FN} \tag{9}$$

where the TPR measures the proportion of actual attacks that are correctly identified, the FPR measures the portion of normal data mistakenly categorized as attacks, and the F-1 score is the harmonic mean of the precision and recall. For the TPR and F-1 score metrics, value approaching 1.0 indicates better performance, while for FPR metric, a value closer to 0 indicates better performance.

To show the effectiveness of the proposed LSTM-Laplace model three baseline models are selected for comparison, which are: LSTM model with Gaussian distribution (LSTM-Gaussian), LSTM model with Gamma distribution (LSTM-Gamma), and quantile regression (QR). The reasons for choosing these baseline models are: (i) QR is a widely used non-parametric probabilistic method [21]. Since the proposed LSTM-Laplace model is a parametric method, the QR baseline allows us to explore the performance between parametric method and non-parametric method; (ii) the LSTM-Gaussian and LSTM-Gamma model allow us to explore the detection performance based on different predictive distribution types.

The evaluation metrics of different models are compared and summarized in Table III. Overall, the proposed LSTM-Laplace anomaly detection method has a higher TPR, F-1 Score, and lower FPR compared with other anomaly detection methods, which shows the effectiveness of the proposed probabilistic anomaly detection algorithm. Note also that the models of LSTM-Gaussian, LSTM-Gamma, and LSTM-Laplace perform similarly and better than the QR method, which indicates that the optimization can help achieve better detection performance with different predictive distribution types in parametric methods. In addition, it is shown that the scores of LRA is better than that of PRA. It is mainly due to the larger LMP magnitude change under LRA and higher PRA attack frequency.

TABLE III: Probabilistic Anomaly Detection Results

Method	Attack Scenario	Metrics		
		TPR	FPR	F-1 Score
LSTM-Laplace	LRA	**0.91**	**0.19**	**0.89**
	PRA	**0.86**	**0.23**	**0.86**
LSTM-Gaussian	LRA	0.89	0.24	0.87
	PRA	0.85	0.24	0.88
LSTM-Gamma	LRA	0.86	0.23	0.87
	PRA	0.85	0.24	0.86
QR	LRA	0.79	0.33	0.84
	PRA	0.71	0.35	0.65

Note: The best TPR, FPR, and F-1 score among different models are marked in boldface.

2) Results Analysis: To better visualize the probabilistic anomaly detection results, the PIs of selected time period under LRA and PRA are illustrated in Fig. 4 and Fig. 5, respectively. It is observed that at most part of the no attack periods, the LMP reasonably lies within the PIs. When the observation in the attack period falls out of the 70% PI , it is defined as a truth positive detection. It is seen that the PRA frequency in Fig. 5 is higher than that of LRA in Fig. 4, and the magnitude change of LMP under LRA is higher than that under PRA. However, under both scenarios, the high detection accuracy shows the robustness of the proposed method. The width of the PI varies with the LMP variability. When the LMP fluctuates more frequently, the PI tends to be wider, and thereby the uncertainty under PRA is relatively higher.

IV. Conclusion

This paper developed a data-driven probabilistic anomaly detection method in electricity market. Results of the case

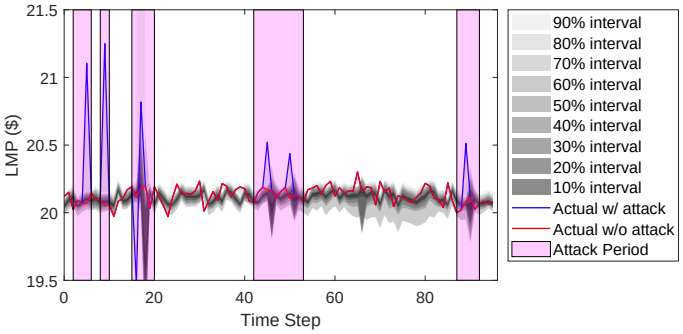

Fig. 4: PIs of LMP under LRA attack

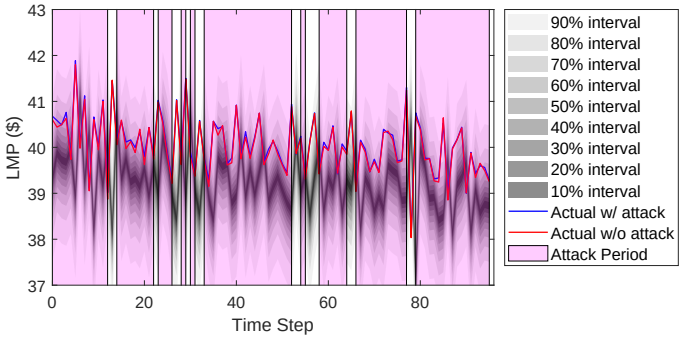

Fig. 5: PIs of LMP under PRA attack

study under different attack scenarios showed that the developed probabilistic anomaly detection method was able to effectively detect both LRA and PRA in electricity market. Future work will explore: (i) performance improvement using spatio-temporal correlations among nearby nodes, and (ii) sensitivity analysis on different PI thresholds to detection performance.

V. Acknowledgement

This material is based upon work supported by the Department of Energy under Award Number DE-OE0000899.

VI. Disclaimer

This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

[1] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[2] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Distributed internet-based load altering attacks against smart power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 667–674, 2011.

[3] S. Amini, F. Pasqualetti, and H. Mohsenian-Rad, "Dynamic load altering attacks against power system stability: Attack models and protection schemes," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2862–2872, 2016.

[4] 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.

[5] A. Dabrowski, J. Ullrich, and E. R. Weippl, "Grid shock: Coordinated load-changing attacks on power grids: The non-smart power grid is vulnerable to cyber attacks as well," in *Proceedings of the 33rd Annual Computer Security Applications Conference*. ACM, 2017, pp. 303–314.

[6] X. Wang, T. Zhao, H. Liu, and R. He, "Power consumption predicting and anomaly detection based on long short-term memory neural network," in *2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE, 2019, pp. 487–491.

[7] V. B. Krishna, G. A. Weaver, and W. H. Sanders, "Pca-based method for detecting integrity attacks on advanced metering infrastructure," in *International Conference on Quantitative Evaluation of Systems*. Springer, 2015, pp. 70–85.

[8] K. G. Lore, D. M. Shila, and L. Ren, "Detecting data integrity attacks on correlated solar farms using multi-layer data driven algorithm," in *2018 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.

[9] S. Bouktif, A. Fiaz, A. Ouni, and M. Serhani, "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, 2018.

[10] C. Olah, "Understanding lstm networks," 2015.

[11] M. Sun, C. Feng, E. K. Chartan, B.-M. Hodge, and J. Zhang, "A two-step short-term probabilistic wind forecasting methodology based on predictive distribution optimization," *Applied energy*, vol. 238, pp. 1497–1505, 2019.

[12] J. R. Koza, "Genetic programming," 1997.

[13] M. Sun, C. Feng, and J. Zhang, "Multi-distribution ensemble probabilistic wind power forecasting," *Renewable Energy*, vol. 148, pp. 135–149, 2020.

[14] C. Feng, M. Sun, and J. Zhang, "Reinforced deterministic and probabilistic load forecasting via q-learning dynamic model selection," *IEEE Transactions on Smart Grid*, 2019.

[15] C. E. M.-S. R. D. Zimmerman, "Matpower (version 7.0)," *[Software] Available at: https://matpower.org*, 2019.

[16] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[17] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: A survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2016.

[18] R. Tan, V. B. Krishna, D. K. Yau, and Z. Kalbarczyk, "Integrity attacks on real-time pricing in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 18, no. 2, pp. 1–33, 2015.

[19] S. Soltan, P. Mittal, and V. Poor, "Protecting the grid against mad attacks," *IEEE Transactions on Network Science and Engineering*, 2019.

[20] M. Sun, C. Feng, and J. Zhang, "Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation," *Applied Energy*, vol. 256, p. 113842, 2019.

[21] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology*, vol. 9, no. 1-2, pp. 95–108, 2006.