Security and Attacks on Federated Energy Forecasting

Jonas Sievers*, Krupali Kumbhani†, Thomas Blank*, Frank Simon*, Andreas Mauthe†

*Karlsruhe Institute of Technology (KIT), Institute for Data Processing and Electronics (IPE), Karlsruhe, Germany, e-mail: jonas.sievers@kit.edu
†University of Koblenz, Institute for Information Systems Research, Koblenz, Germany

***Abstract*—Accurate energy forecasting, including load, photovoltaic generation, and prosumption prediction, is essential for the efficient operation and strategic planning of modern energy systems. Federated Learning (FL) has emerged as a promising solution for training machine learning models on decentralized data, enabling high model accuracy while maintaining data privacy. However, the decentralized nature of FL also poses security challenges, including data poisoning and backdoor attacks that compromise the integrity and reliability of forecasting models. In this study, we present a comprehensive evaluation of various data poisoning and backdoor attacks within federated energy forecasting. Our analysis explores different data distributions, varying noise scales in data poisoning attacks, and targeted manipulation of specific time intervals to assess their impact on model performance. Further, we propose robust security mechanisms, such as increased cluster sizes, local retraining, and weighted aggregation. Our results show that while our attacks can increase the Mean Absolute Error by 93-261 %, our security measures can effectively mitigate the attacks, thereby improving the security and robustness of federated energy forecasting.**

Keywords*- *federated learning; poisoning attack; backdoor attack.

I. Introduction

The transition to sustainable energy systems is essential for addressing climate change and reducing the dependence on fossil fuels. As countries decarbonize their grids, accurate energy forecasting becomes critical to balance renewable energy supply and demand [1]. Reliable grid operation depends on accurate predictions of electric loads, photovoltaic (PV) generation, and prosumption patterns, especially as energy grids become more decentralized and complex [2].

Here, Federated Learning (FL) has been proposed for energy forecasting, enhancing model performance, data efficiency, and privacy. As shown in Figure 1, only model parameters are shared with a central server in FL, while local data remains private [3]. This approach minimizes the risk of exposing sensitive consumption patterns, which could otherwise be exploited to infer personal habits, posing privacy threats [4]. Additionally, clustering is applied in FL to group nodes with similar energy patterns, addressing challenges with non independent and identically distributed (non-iid) datasets.

While FL enhances privacy and security, challenges such as data poisoning and backdoor attacks persist. Data poisoning skews model performance by manipulating local data, while backdoor attacks insert hidden triggers into the model that only activate malicious behavior under specific conditions. While these vulnerabilities have been studied within the vision and language domains [5], they are particularly concerning in energy systems, where forecasting errors can disrupt supply and demand balance, hindering renewable energy integration [6]. Addressing these risks is crucial to improve the robustness of FL in energy forecasting.

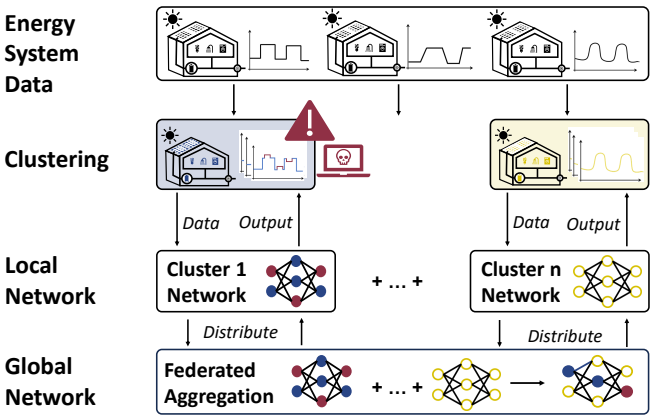

Figure 1. Clustered federated learning architecture with data poisoning.

To address security challenges in FL-based energy forecasting applications, we investigate the effects of data poisoning and backdoor attacks. By manipulating data at varying scales and time periods, we assess model vulnerabilities, identifying weaknesses in the FL framework. We then propose defense strategies, including secure aggregation, local retraining, and clustering methods, to enhance the system's resilience against adversarial threats.

A. Related Work

To provide a comprehensive understanding of the current research and challenges in federated energy systems, we review related work on security measures and adversarial attacks. Additionally, we examine implementations of these attacks in the domains of computer vision and natural language processing. Selected publications are summarized in Table I.

McMahan et al. introduced FL in 2016 [3] as a method for decentralized model training across distributed devices, preserving data privacy by keeping data localized. While FL has been applied in the energy domain for applications like energy control [20]–[22], non-intrusive load monitoring [23]–[25], and energy theft detection [26], research on adversarial attacks within federated energy systems remains limited. Here, only

TABLE I
REVIEW OF SECURITY AND ATTACK LITERATURE IN FEDERATED ENERGY FORECASTING.

Ref	Year	Focus	Domain	Attack	Security
[7]	2021	Differential Privacy	Energy		✓
[8]	2023	Differential Privacy	Energy		✓
[9]	2022	Differential Privacy	Energy		✓
[10]	2024	Secure Aggregation	Energy		✓
[11]	2023	Secure Aggregation	Energy		✓
[12]	2023	Secure Aggregation	Energy		✓
[13]	2023	Secure Aggregation	Energy		✓
[14]	2022	Secure Aggregation	Energy		✓
[15]	2023	Secure Aggregation	Energy		✓
[16]	2023	Personalized FL	Energy		✓
[17]	2022	Model Poisoning	Energy	✓	
[18]	2020	Inference Attacks	Vision	✓	
[19]	2019	Poisoning Attacks	Language	✓	
This paper	2024	Attacks and Security	Energy	✓	✓

model poisoning has been analyzed [17], as most research focuses on security measures [7]–[16]. In contrast, adversarial attacks – including model poisoning, inference attacks, data poisoning, and backdoor attacks – have been extensively studied in domains like computer vision [18], [27], [28] and natural language processing [19], [29], [30], highlighting significant risks to FL models.

B. Paper Contribution and Organization

Vulnerabilities specific to federated energy forecasting have so far not been thoroughly investigated. To address this gap, we analyze adversarial attacks in federated energy forecasting and propose mitigation strategies. Since most research on FL attacks focuses on natural language processing and computer vision, applying those findings to energy prediction is challenging due to different data characteristics and dimensionality. Consequently, our main contributions are:

- We develop data poisoning and backdoor attacks customized for energy forecasting, evaluating their impact on model performance in FL systems using selected noise distributions (Uniform, Normal, Laplace, Building's) and targeting specific time intervals.
- We benchmark these attacks across different model architectures, including a Bidirectional Long-Short Term Memory Model (BiLSTM), a Soft-Gated LSTM (Soft-LSTM), and a Soft-Gated Dense Neural Network (Soft-Dense).
- We integrate security mechanisms such as secure aggregation, varying cluster sizes, and local retraining, to mitigate the effects of these attacks.
- Our findings show that data poisoning attacks significantly impact model performance, especially in small clusters, while backdoor attacks pose minor threats. By incorporating our proposed security measures, these adverse effects are mitigated, enhancing the security and robustness of FL-based energy forecasting systems.

The reminder of the paper is organized as follows: Section II introduces our methodology, while Section III outlines our experimental setup. Building on this, Section IV presents our results, Section V discusses our results and limitations, and Section VI provides our conclusion and future work.

II. METHODOLOGY

In this section, we provide a concise overview of our methodology, including federated energy systems, data poisoning, backdoor attacks and security measures.

A. Federated Energy Systems

In federated energy systems, a central server initializes global model weights w_0 for each cluster and distributes them to local devices (clients). Each client i trains a local model on its local dataset D_i and returns updated weights w_i. The server aggregates these weights using a selected aggregation method. One common approach is Average Aggregation, where the global model weights w_{global} are updated as:

$$w_{global} = \frac{1}{N} \sum_{i=1}^{N} w_i \tag{1}$$

where N is the number of clients within a cluster. This process is repeated over t federated training rounds.

B. Data Poisoning Attack in Federated Energy Systems

Data poisoning attacks threaten federated energy systems by compromising the integrity of the global FL model. Attackers manipulate the local datasets of specific clients, distorting energy forecasts and leading to inaccurate predictions that can affect operations like load balancing or grid management.

For a data point (x, y), with the input vector x and the target vector y, the poisoned input vector x' is defined as:

$$x' = x + \epsilon, \quad \epsilon \sim \mathcal{D}(\gamma) \tag{2}$$

Here, ϵ represents noise sampled from a distribution $\mathcal{D}$ with noise scale γ. Possible choices for $\mathcal{D}$ include Normal, Laplace, and Uniform distributions, or the distribution of the actual building measurements. To ensure proportionate noise injection across varying energy data scales, we normalize x to the range $[0, 1]$. During local training, attacked clients $j \in \mathcal{A}$ use the poisoned data x'_j, while benign clients $k \in \mathcal{B}$ use unmodified data. The weight update for benign clients is:

$$\mathbf{w}_{k,\text{new}} = \mathbf{w}_k - \eta \frac{\partial \mathcal{L}(\mathbf{w}_k, \mathbf{x}_k, y_k)}{\partial \mathbf{w}_k} \tag{3}$$

For attacked clients $j \in \mathcal{A}$, the weight update is:

$$\mathbf{w}'_{j,\text{new}} = \mathbf{w}_j - \eta \frac{\partial \mathcal{L}(\mathbf{w}_j, \mathbf{x}'_j, y_j)}{\partial \mathbf{w}_j} \tag{4}$$

Here, η is the learning rate, and $\frac{\partial \mathcal{L}}{\partial \mathbf{w}}$ denotes the gradient of the loss function $\mathcal{L}$ with respect to the model weights $\mathbf{w}$.

During the federated aggregation, the global model weights $\mathbf{w}'_{\text{global}}$ are updated by averaging:

$$\mathbf{w}'_{\text{global}} = \frac{1}{N} \left(\sum_{k \in \mathcal{B}} \mathbf{w}_{k,\text{new}} + \sum_{j \in \mathcal{A}} \mathbf{w}'_{j,\text{new}} \right) \tag{5}$$

where N is the total number of clients. The inclusion of poisoned weights $\mathbf{w}'_{j,\text{new}}$ can degrade global model performance by introducing biased patterns. Systematically testing different distributions $\mathcal{D}$ and noise scales γ allows us to evaluate the model's vulnerability to these attacks.

C. Backdoor Attack in Federated Energy Systems

Backdoor attacks in federated energy systems manipulate client data during specific hours, potentially causing the FL model to produce inaccurate forecasts during peak hours while maintaining normal performance at other times.

Therefore, attackers adjust the input data only during selected hours $H \subseteq 0, 1, \ldots, 23$. For each data point (t_j, x_j, y_j), where t_j denotes the hour, the modified input feature vector x'_j is defined as:

$$x'_j = \begin{cases} x_j + \delta, & \text{if } t_j \in H \\ x_j, & \text{otherwise} \end{cases} \tag{6}$$

Here, δ represents the backdoor trigger – a specific perturbation added only during the targeted hours H. The local training and federated aggregation remain the same as described in Subsection II-B. By adjusting δ and selecting specific hours H, attackers can fine-tune the severity of the backdoor attack and evaluate the model's performance.

D. Security Measures in Federated Energy Systems

To mitigate data poisoning and backdoor attacks in federated energy systems, we propose three security strategies: clustering, weighted aggregation, and local retraining.

Clustering reduces the impact of attacks by grouping clients with similar time series, restricting the extent of manipulation before a client is excluded from the cluster. Clients are grouped together when their time series E and F satisfy the similarity condition $d(E, F) \leq \tau$. A common similarity measure is Dynamic Time Warping (DTW), which minimizes the cumulative distance over all possible alignments (m_i, n_i):

$$d_{\text{DTW}}(E, F) = \sqrt{\min_{m_i, n_i} \left(\sum_{i=1}^{I} (e_{m_i} - f_{n_i})^2 \right)}, \tag{7}$$

Selecting an appropriate τ ensures that deviations from attacks remain within acceptable bounds.

Weighted Aggregation mitigates the attack effects by adjusting client contributions based on local model performance. Clients exhibiting degraded performance due to attacks receive lower weights, reducing their influence on the global model. Given a sets of benign clients $\mathcal{B}$ and attacked clients $\mathcal{A}$, the aggregation is performed as follows:

$$\mathbf{w}_{\text{global}} = \frac{\sum_{k \in \mathcal{B}} \alpha_k \cdot \mathbf{w}_k + \sum_{j \in \mathcal{A}} \alpha'_j \cdot \mathbf{w}'_j}{\sum_{k \in \mathcal{B}} \alpha_k + \sum_{j \in \mathcal{A}} \alpha'_j} \tag{8}$$

where α_k is the weight for benign clients, and α'_j represents the weight for attacked clients, typically $\alpha'_j \ll \alpha_k$.

Local Retraining allows benign clients to adapt the global model to their local data, reducing the impact of poisoned global weights w'_{global}. After receiving $\mathbf{w}'_{\text{global}}$, benign clients refine their models:

$$\mathbf{w}_{k,\text{retrained}} = \mathbf{w}_{\text{global}} - \eta \frac{\partial \mathcal{L}(\mathbf{w}_{\text{global}}, \mathbf{x}_k, y_k)}{\partial \mathbf{w}_{global}} \tag{9}$$

where η is the learning rate, $\mathcal{L}$ is the loss function, and $(\mathbf{x}_k, y_k)$ represents the local dataset. This fine-tuning mitigates attack influence and enhances model robustness.

Together, these strategies enhance the security and reliability of federated models in energy systems by effectively countering adversarial attacks.

III. Experimental Setup

Building on our methodology, we describe our experimental setup, including data analysis and federated energy forecasting.

A. Data Analysis

We utilize the Ausgrid dataset [31], which provides half-hourly smart meter readings of electrical load and PV output in kW from 300 residential buildings in Australia between 2010 and 2013. An example of load and PV patterns for Building 11 is shown Figure 2.

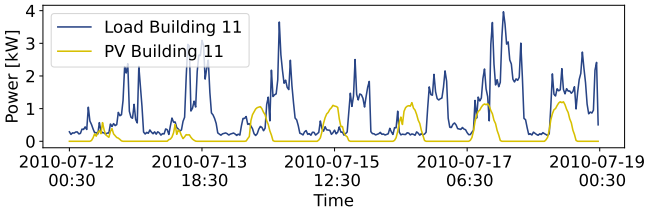

Figure 2. Load and PV patterns of Building 11.

We extend the dataset by calculating prosumption (load - PV). For computational efficiency, our analysis focuses on a randomly selected subset of the first 30 households. To obtain our forecasting dataset, we use 70% for training, 20% for validation, and 10% for testing.

B. Federated Energy Forecasting

Our FL architecture consists of 3 training rounds, as additional rounds did not yield further improvements. In each round, clients update their local models and send them to a central server for global aggregation. We use K-Means clustering with DTW to group clients with similar energy patterns into 10 clusters. For testing adversarial attacks, we focus on a cluster containing 2 clients (buildings 16 and 24), simulating distributed training on a single machine.

To implement federated energy forecasting, we employ a BiLSTM network and two Mixture of Experts (MoE) models within the FL framework. The MoE architecture enhances the model's ability to learn complex patterns by dynamically selecting and weighting outputs from multiple specialized sub-models (experts) based on the input sequences. Specifically, the Soft-Dense model includes an expert layer with four experts (each with eight units), followed by two Dense layers (16 units each), a Dropout layer (rate 0.2), and a Flatten layer.

The Soft-LSTM model comprises an expert layer (four experts, eight units each), a bidirectional LSTM layer (four units), a Dropout layer (rate 0.2), and a Flatten layer. The BiLSTM network captures temporal dependencies in energy data through bidirectional processing, using two layers of eight LSTM cells each. All models use a batch size of 16. Further architectural details are provided in [4].

To evaluate robustness against adversarial attacks, we implement data poisoning and backdoor scenarios. In the data poisoning attack, noise is injected into the data from various distributions. For Normal and Laplace distributions, the mean is 0, and the standard deviation varies from 0 to 1. For the Uniform distribution, bounds range between $[-1, 0]$ and $[0, 1]$. Building-specific distributions are derived based on skewness, kurtosis, and mean, with standard deviations varying between 0 and 1. In the backdoor attack, we manipulate four specific half-hour time steps during hours 0 and 1, setting load and prosumption values to zero and PV values to one.

As security measures, we expand the cluster size to 4 buildings and implement a local retraining step with 100 epochs, incorporating early stopping if the validation loss does not improve over 10 consecutive epochs. Weighted aggregation is not employed in this experiment, as clustering and retraining already mitigate the effects. For performance evaluation, we calculate the Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{10}$$

This metric quantifies the average absolute difference between the predicted values $\hat{y}_i$ and actual values y_i.

IV. Results

In this section, we present our results for data poisoning, backdoor attacks, and security measures. Within each attack scenario, the *poisoned* model is trained with manipulated data, while the *unmodified* model is only indirectly affected through FL. To consider statistical variations, each model is trained 3 times per scenario. If not stated otherwise, the indicated metrics are averaged over all buildings, clusters, or training rounds and the results are achieved on the test dataset.

A. Data Poisoning in Federated Energy Forecasting

Within the data poisoning attack, noise patterns are introduced to the training data using four distributions: Normal, Laplace, Uniform, and building-specific. Due to space constraints, we report results only for the Uniform distribution, which had the most significant effects.

Figure 3 shows the MAE for both poisoned models (top row) and unmodified models (bottom row) with increasing noise scales over load, PV, and prosumption forecasting. The dashed lines represent the baseline performance of unmodified models within local learning, to evaluate whether the attacked FL architecture still provides performance benefits.

The results indicate a substantial increase in MAE with rising noise scale for all poisoned models. Specifically, the Soft-Dense model exhibits increases in MAE of 321 % for load, 718 % for PV, and 345 % for prosumption forecasting. Similar trends are observed in the BiLSTM and Soft-LSTM models. It is worth noting that as the MAE values are small, minor increases already result in large percentage changes. The unmodified models, indirectly affected through FL, also showes significant degradation, with MAE increases of up to 96% for load, 93% for PV, and 261% for prosumption forecasting in the Soft-Dense model.

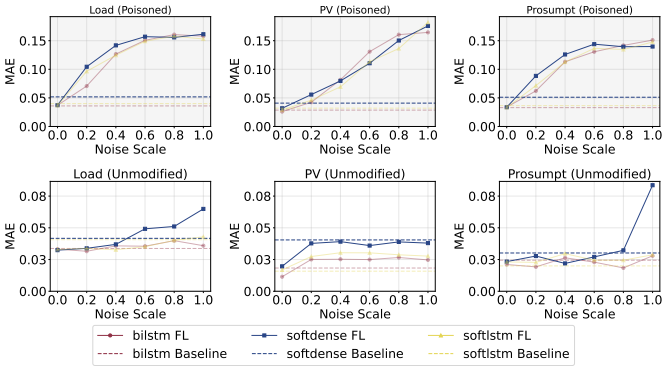

Figure 3. Performance comparison between poisoned and unmodified models.

Without poisoning, FL generally outperforms local learning. However, when subject to attack, the unmodified Soft-Dense model's performance droped below the local learning baseline for load forecasting at a noise scale of 0.6 and for prosumption prediction at 1.0. In contrast, PV forecasting performance remained close to the local baseline across all noise scales. Detailed results for the unmodified Soft-Dense model are provided in Table II.

TABLE II
Model performance of the unmodified Soft-Dense model for Local Learning (LL) and FL with different noise Scales (N), where N0.2 corresponds to a noise scale of 0.2.

Noise	Load		PV		Prosumption	
	MAE	STD	MAE	STD	MAE	STD
LL	0.0417	±0.0051	0.0405	±0.0055	0.0302	±0.0068
N0	0.0324	±0.0196	0.0196	±0.0164	0.0234	±0.0140
N0.2	0.0338	±0.0047	0.0377	±0.0248	0.0278	±0.0023
N0.4	0.0369	±0.0016	0.0391	±0.0258	0.0219	±0.0017
N0.6	0.0492	±0.0007	0.0360	±0.0265	0.0271	±0.0029
N0.8	0.0511	±0.0028	0.0389	±0.0268	0.0323	±0.0079
N1	0.0649	±0.0255	0.0379	±0.0261	0.0836	±0.0012

B. Backdoor Attack in Federated Energy Forecasting

In the backdoor attack scenario, data is selectively modified for specific hours, using the date as the trigger. Figure 4 shows the MAE for each hour of the day for both poisoned models (top row) and unmodified models (bottom row). Within each subplot, solid lines represent the performance of the FL architectures affected by the attack, while dashed lines indicate the baseline performance without any attack. The hourly MAE naturally fluctuates due to inherent volatility variations.

For the Soft-Dense model with the backdoor attack, the MAE increased significantly only in prosumption forecasting, rising

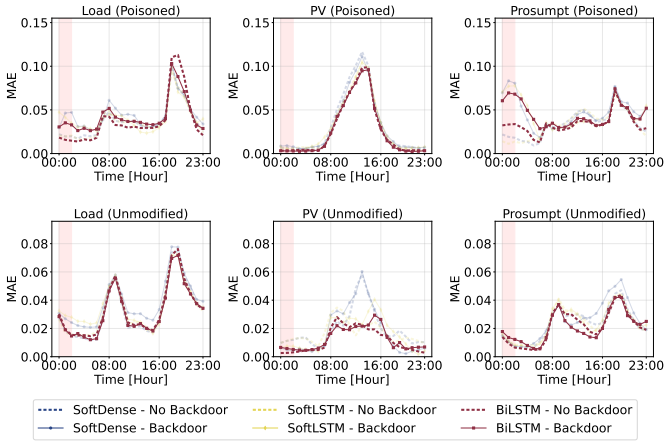

Figure 4. Model performance comparison with and without backdoor attacks.

by 0.0560 (273%). Changes in load (0.0142, 68%) and PV (0.0022, 33%) forecasting remain within the standard deviation. The unmodified models show minimal changes, with slight increases in load (0.0083, 36%) and prosumption (0.0001, 1%), and a decrease in PV (-0.0051, -49%), all within the standard deviation. Detailed MAE values of the Soft-Dense model for the attacked hours are provided in Table III, comparing the backdoor model to the baseline.

TABLE III
PERFORMANCE METRICS OF THE FORECASTING MODELS WITH AND WITHOUT BACKDOOR ATTACKS.

Scenario	Load		PV		Prosumpt.	
	MAE	STD	MAE	STD	MAE	STD
Pois. Back.	0.0351	0.0167	0.0089	0.0064	0.0765	0.0164
Pois. noBack.	0.0209	0.0110	0.0067	0.0017	0.0205	0.0105
Pois. Diff.	0.0142	0.0057	0.0022	0.0047	0.0560	0.0059
Un. Back.	0.0312	0.0073	0.0054	0.0023	0.0127	0.0039
Un. noBack.	0.0229	0.0063	0.0105	0.0060	0.0126	0.0025
Un. Diff.	0.0083	0.0010	-0.0051	-0.0037	0.0001	0.0014

C. Security in Federated Energy Forecasting

We mitigate attacks by increasing the cluster size, reducing the influence of compromised buildings on the aggregation model, while local retraining refines model parameters using unmodified data. Due to space constraints, Figure 5 illustrates that the MAE remains stable for all unmodified models and noise scales, indicating that this combination effectively mitigates the attacks.

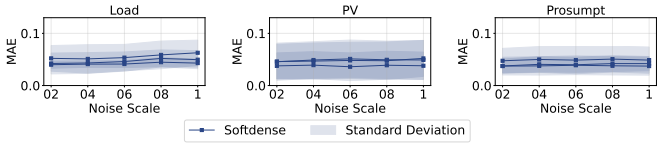

Figure 5. Impact of security measures on the unmodified forecasting models.

While these measures are effective, incorporating weighted average aggregation could further improve resilience, by weighting models based on their performance, thus reducing the contribution of compromised models.

V. DISCUSSION AND LIMITATIONS

Our data poisoning attacks substantially increased the MAE across all models, primarily due to the limited cluster sizes that increase the impact of compromised data. Load forecasting, due to its inherently stochastic nature, experienced greater performance degradation, whereas PV prediction maintained more stable performance, benefiting from its more deterministic patterns. Conversely, backdoor attacks had a minimal overall effect, as their targeted manipulations were limited to specific forecasting times, thereby reducing their influence on the aggregated model. The implemented security measures – namely, increasing cluster size and applying local retraining – effectively mitigate these attacks by reducing the impact of poisoned data. In addition, incorporating weighted average aggregation could further enhance resilience by reducing the contribution of compromised models during federated aggregation.

This study is limited by its focus on only few buildings, which may affect the generalizability of our findings. Incorporating a broader range of benchmark models and datasets could further validate the robustness of the proposed defense mechanisms. Further, we use simple noise sampling methods for data poisoning, which may not fully capture the complexity of more advanced attacks.

VI. CONCLUSION AND FUTURE WORK

In this paper, we comprehensively analyzed security vulnerabilities in federated energy forecasting, focusing on the impacts of data poisoning and backdoor attacks. Our findings demonstrate that data poisoning poses a significant threat to forecasting accuracy, with MAE increasing by 93-261 %, especially within smaller clusters. Conversely, backdoor attacks show a limited impact on model performance. By incorporating defense mechanisms such as increased cluster sizes and local retraining, we effectively enhanced the resilience of federated learning models, mitigating the adversarial risks and preserving model integrity. Future work could explore the use of Generative Adversarial Networks for sophisticated noise generation during training.

REFERENCES

[1] F. Plaum, R. Ahmadiahangar, A. Rosin, and J. Kilter, "Aggregated demand-side energy flexibility: A comprehensive review on characterization, forecasting and market prospects", *Energy Reports*, vol. 8, pp. 9344–9362, 2022, ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2022.07.038.

[2] U. Das *et al.*, "Forecasting of photovoltaic power generation and model optimization: A review", *Renewable Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018. DOI: 10.1016/J.RSER.2017.08.017.

[3] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data", in *International Conference on Artificial Intelligence and Statistics*, 2016.

[4] J. Sievers, T. Blank, and F. Simon, "Advancing Accuracy in Energy Forecasting using Mixture-of-Experts and Federated Learning", in *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*, ser. e-Energy '24, Singapore, Singapore: Association for Computing Machinery, 2024, pp. 65–83, ISBN: 9798400704802. DOI: 10.1145/3632775.3661945.

[5] J. Zhang *et al.*, "Security and Privacy Threats to Federated Learning: Issues, Methods, and Challenges", *Sec. and Commun. Netw.*, vol. 2022, 2022, ISSN: 1939-0114. DOI: 10.1155/2022/2886795.

[6] J. Jasiūnas, P. D. Lund, and J. Mikkola, "Energy system resilience – A review", *Renewable and Sustainable Energy Reviews*, vol. 150, p. 111 476, 2021, ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2021.111476.

[7] Y. Zhao *et al.*, "A Differential Privacy-enhanced Federated Learning Method for Short-Term Household Load Forecasting in Smart Grid", Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1399–1404, ISBN: 9781665409506. DOI: 10.1109/ICCC54389.2021.9674514.

[8] X. Qu *et al.*, "Personalized Federated Learning for Heterogeneous Residential Load Forecasting", *Big Data Mining and Analytics*, vol. 6, pp. 421–432, 4 2023, ISSN: 20960654. DOI: 10.26599/BDMA.2022.9020043.

[9] M. A. Husnoo *et al.*, "A Secure Federated Learning Framework for Residential Short Term Load Forecasting", 2022.

[10] Y. Dong *et al.*, "Privacy-Preserving Distributed Learning for Residential Short-Term Load Forecasting", *IEEE Internet of Things Journal*, 2024, ISSN: 23274662. DOI: 10.1109/JIOT.2024.3362587.

[11] J. Li, H. Li, R. Wang, Y. Guo, and S. Wu, "Fed-SAD:A secure aggregation federated learning method for distributed load forecasting", 2023. DOI: 10.22541/au.169028986.64063960/v1.

[12] M. A. Husnoo *et al.*, "FedDiSC: A computation-efficient federated learning framework for power systems disturbance and cyber attack discrimination", *Energy and AI*, vol. 14, 2023, ISSN: 26665468. DOI: 10.1016/j.egyai.2023.100271.

[13] Y. Liu, Z. Dong, B. Liu, Y. Xu, and Z. Ding, "FedForecast: A federated learning framework for short-term probabilistic individual load forecasting in smart grid", *International Journal of Electrical Power and Energy Systems*, vol. 152, 2023, ISSN: 01420615. DOI: 10.1016/j.ijepes.2023.109172.

[14] M. M. Badr *et al.*, "Privacy-Preserving Federated-Learning-Based Net-Energy Forecasting", vol. 2022-March, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 133–139, ISBN: 9781665406529. DOI: 10.1109/SoutheastCon48659.2022.9764093.

[15] H. U. Manzoor, A. R. Khan, T. Sher, W. Ahmad, and A. Zoha, "Defending Federated Learning from Backdoor Attacks: Anomaly-Aware FedAVG with Layer-Based Aggregation", Institute of Electrical and Electronics Engineers Inc., 2023, ISBN: 9781665464833. DOI: 10.1109/PIMRC56721.2023.10293950.

[16] F. Widmer, S. Nowak, B. Bowler, P. Huber, and A. Papaemmanouil, "Data-driven comparison of federated learning and model personalization for electric load forecasting", *Energy and AI*, vol. 14, 2023, ISSN: 26665468. DOI: 10.1016/j.egyai.2023.100253.

[17] N. B. S. Qureshi, D. H. Kim, J. Lee, and E. K. Lee, "Poisoning Attacks against Federated Learning in Load Forecasting of Smart Energy", Institute of Electrical and Electronics Engineers Inc., 2022, ISBN: 9781665406017. DOI: 10.1109/NOMS54207.2022.9789884.

[18] X. Luo and X. Zhu, "Exploiting defenses against gan-based feature inference attacks in federated learning", *CoRR*, vol. abs/2004.12571, 2020.

[19] J. Zhang, J. Chen, D. Wu, B. Chen, and S. Yu, "Poisoning attack in federated learning using generative adversarial nets", Institute of Electrical and Electronics Engineers Inc., 2019, pp. 374–380, ISBN: 9781728127767. DOI: 10.1109/TrustCom/BigDataSE.2019.00057.

[20] S. Lee, L. Xie, and D.-H. Choi, "Privacy-Preserving Energy Management of a Shared Energy Storage System for Smart Buildings: A Federated Deep Reinforcement Learning Approach", *Sensors*, vol. 21, no. 14, 2021, ISSN: 1424-8220. DOI: 10.3390/s21144898.

[21] S. Lee and D.-H. Choi, "Federated Reinforcement Learning for Energy Management of Multiple Smart Homes With Distributed Energy Resources", *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 488–497, 2022. DOI: 10.1109/TII.2020.3035451.

[22] F. Rezazadeh and N. Bartzoudis, "A Federated DRL Approach for Smart Micro-Grid Energy Control with Distributed Energy Resources", 2022. DOI: 10.1109/CAMAD55695.2022.9966919.

[23] A. Giuseppi, S. Manfredi, D. Menegatti, A. Pietrabissa, and C. Poli, "Decentralized Federated Learning for Nonintrusive Load Monitoring in Smart Energy Communities", in *2022 30th Mediterranean Conference on Control and Automation (MED)*, 2022, pp. 312–317. DOI: 10.1109/MED54222.2022.9837291.

[24] Y. Wang, I. L. Bennani, X. Liu, M. Sun, and Y. Zhou, "Electricity Consumer Characteristics Identification: A Federated Learning Approach", *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3637–3647, 2021. DOI: 10.1109/TSG.2021.3066577.

[25] Y. He, F. Luo, G. Ranzi, and W. Kong, "Short-Term Residential Load Forecasting Based on Federated Learning and Load Clustering", in *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2021, pp. 77–82. DOI: 10.1109/SmartGridComm51999.2021.9632314.

[26] M. M. Ashraf *et al.*, "FedDP: A Privacy-Protecting Theft Detection Scheme in Smart Grids Using Federated Learning", *Energies*, vol. 15, no. 17, 2022, ISSN: 1996-1073. DOI: 10.3390/en15176241.

[27] R. Mayerhofer and R. Mayer, "Poisoning Attacks against Feature-Based Image Classification", in *Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy*, ser. CODASPY '22, Baltimore, MD, USA: Association for Computing Machinery, 2022, pp. 358–360, ISBN: 9781450392204. DOI: 10.1145/3508398.3519363.

[28] Z. Xiang, D. J. Miller, and G. Kesidis, "A Benchmark Study Of Backdoor Data Poisoning Defenses For Deep Neural Network Classifiers And A Novel Defense", in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6. DOI: 10.1109/MLSP.2019.8918908.

[29] S. Zhai *et al.*, "Text-to-Image Diffusion Models can be Easily Backdoored through Multimodal Data Poisoning", in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23, Ottawa ON, Canada: Association for Computing Machinery, 2023, pp. 1577–1587, ISBN: 9798400701085. DOI: 10.1145/3581783.3612108.

[30] A. Wan, E. Wallace, S. Shen, and D. Klein, "Poisoning Language Models During Instruction Tuning", in *Proceedings of the 40th International Conference on Machine Learning*, A. Krause *et al.*, Eds., ser. Proceedings of Machine Learning Research, vol. 202, PMLR, 2023, pp. 35 413–35 425.

[31] Elizabeth L. Ratnam, Steven R. Weller, Christopher M. Kellett and Alan T. Murray, "Residential load and rooftop PV generation: an Australian distribution network dataset", *International Journal of Sustainable Energy*, vol. 36, no. 8, pp. 787–806, 2017. DOI: 10.1080/14786451.2015.1100196.