Adversarial Attacks on Demand Side Management of a Grid-Scale Battery Storage

Eike Cramer a\*

a *RWTH Aachen University – Process Systems Engineering (AVT.SVT), Forckenbeckstraße 51, 52074 Aachen, Germany*

E-Mail: eike.cramer@alumni.tu-berlin.de

Abstract

This work investigates the effects of adversarial attacks on the combined decision-making process of electricity price forecasting and optimization-based demand side management (DSM). At the example of a grid-scale battery, this work shows how attackers can induce significant changes in system operation by adding targeted modifications to the input data of the decision-making process. Furthermore, this work proposes a black-box approach using empirical emulators (adversarial surrogate models) of the decision-making process. The proposed attack leads to significant changes in the DSM schedules and the obtained profits, even for small perturbations of the input data.

**Keywords**: Demand side management, electricity price forecasting, machine learning, adversarial attacks, safety

* 1. Introduction

The process systems engineering (PSE) community is and has been at the forefront of developing algorithms for optimization-based DSM for flexible process operation, which can achieve cost savings in the energy- and chemical industries (Ave et al., 2018; Baader et al., 2020). Solving DSM problems typically requires solving large-scale optimization problems, which use predictions of electricity price data as parameters (Morales et al., 2014).

The success and increased application of machine learning-based decision making in DSM and other PSE disciplines increases the risk of malicious interference from outside parties. Most machine learning algorithms used in DSM do not consider the threat of adversarial attacks, i.e., data corruption aimed at deteriorating machine learning model outputs (Xu et al., 2020). This work investigates the vulnerability of flexibly operated industrial processes to such adversarial attacks. Opposed to the known threat of hackers, adversarial attacks are data-based attacks that implement attacks without access to the operation itself. Attack design methods, such as the fast gradient sign method (FGSM) (Goodfellow et al., 2015), rely on the sensitivity information of the machine learning model with respect to its input features to generate adversarial noise that is added to the original data. Such adversarial noise patterns can either pursue a targeted attack, i.e., aiming to induce specific predictions or untargeted attacks, i.e., forcing predictions that are as far from the correct output as possible (Xu et al., 2020).

At the example of a grid-scale battery, this work showcases the effects of adversarial attacks on the decisions made in optimization-based DSM. The grid battery operators aim to return a profit by trading on the day-ahead electricity market. As the actual prices are unknown, the full decision-making process consists of electricity price forecasting and solving a DSM optimization problem. As in most real-world applications, neither the forecasting model nor the DSM optimization model are publicly available. Therefore, this work proposes to train an emulating regression model of the full decision-making process of electricity price forecasting and optimization via a black-box attack scheme. This type of process emulating model is referred to as adversarial surrogate model (ASM) throughout this work. Using gradients of the ASM, the attacker designs minimal perturbations to the residual load forecasts that aim to deteriorate the trading decisions and lead to financial losses.

The trading decisions of the battery operators are made using a linear scheduling model. The optimization problem is solved using electricity price forecasts from state-of-the-art forecasting models such as LASSO regression and artificial neural networks (ANN) that use residual load forecasts as their input features (Trebbien et al., 2023). The results show how minimal changes in the input data can induce significant financial losses to the operation of the electricity storage. Thus, data-based adversarial attacks pose a threat to comparable decision-making processes in the energy- and chemical industries.

* 1. DSM of a grid-scale battery and multi-period electricity price forecasting

This Section introduces the DSM case study and the multi-period electricity price forecasting scheme. Note that this Section states the models for completeness. The attacker has no access to of knowledge of the scheduling or the forecasting model.

* + 1. DSM of a Battery Storage

This work investigates the DSM of a grid-scale battery. The battery operation aims to return a profit by trading on the day-ahead market. Problem (P) shows the optimization problem to determine the optimal trading decisions. The problem considers a 1200 kWh storage capacity, and the formulation is adapted from previous work (Cramer et al., 2022) and general formulations for day-ahead trading (Morales et al., 2014). The formulation assumes a constant efficiency for charging and discharging, making Problem (P) a linear optimization problem.

|  |  |  |
| --- | --- | --- |
|  | | (P) |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

In Problem (P), and are the trading decisions for the -th hour, is the maximum (dis-)charging rate, is the state of charge at hour , is the day-ahead price at the hour , and is the trading interval of 1h. Generally, the battery will charge during low-price hours and discharge during high-price hours. The problem solves efficiently using the *gurobi* optimization software.

* + 1. Multi-period electricity price forecasting

The decision-making process uses multi-period forecasting to predict the electricity price based on day-ahead residual load forecasts, which are the most important features for day-ahead electricity prices (Trebbien et al., 2023). The multi-period approach matches the structure of the day-ahead markets where all 24-time intervals are set at the same time. The general form of the multi-period forecasting is a multivariate regression problem that reads:

|  |  |
| --- | --- |
|  | (1) |

Here, is the vector of day-ahead electricity prices, are the residual load forecasts, and is a multivariate regression model. To implement the multivariate regression, this work uses a linear regression model with a LASSO penalty, a fully connected ANN, and a convolutional ANN (TCN). All regression models are implemented using the Python-based machine learning library TensorFlow.

* 1. Black-box attacks using adversarial surrogate models

This Section proposes an approach to fitting an emulator to the full decision-making process for DSM. The ASM is used to compute sensitivities that provide quantitative information to design adversarial noise.

* + 1. Ein Bild, das Screenshot enthält.

       Automatisch generierte BeschreibungEmulating decision-making processes via adversarial surrogate models

Figure 1 Concept of the adversarial surrogate model (ASM). EPF and DSM models are inside the secure company boundary. The ASM is trained to emulate the full decision-making process.

The individual steps of the decision-making process are unknown to external attackers, i.e., the attacker does not know the EPF model nor the DSM model. Furthermore, the EPF takes place within the secure company network and, thus, the prices forecasts cannot be intercepted by the attacker. Instead, the attacker only knows the input and output data. For the considered case, this assumption is reasonable as the final DSM decisions are the target values of the attacks and the input data consists of the most widely used input data for electricity prices forecasting, i.e., the residual load forecasts (c.f. Section 2.2 for details). This work proposes a black-box attack strategy based on what this work calls an adversarial surrogate model (ASM). Thus, the attack only has to intervene with the data pipelines to and from the company instead of the infiltrating the company network itself.

Figure 1 shows the ASM scheme. The ASM is trained on the residual load forecasts and the buy/sell commitments of the company under attack. The evaluation of the ASM then reads:

|  |  |
| --- | --- |
|  | (2) |

Here, are the ASM prediction of the trading decisions. In the following, the ASM provides gradient information that describes how the decisions of the company change based on changes to the input data, i.e., the residual load forecasts.

* + 1. Heuristic attack target for the design of adversarial attacks

Adversarial attacks allow for either targeted or untargeted attacks (Xu et al., 2020), i.e., attacks that pursue a specific target or randomly worsen the outputs, respectively. This work proposes a dampening heuristic to design targeted attacks. Following the intuition that a constant operation is worse than a flexible operation based on variable electricity prices, the proposed dampening attack aims to force trading decisions towards a constant.

This work uses the simple yet effective fast gradient sign method (FGSM) (Goodfellow et al., 2015) to compute adversarial noise patterns that are added to the original input data:

|  |  |
| --- | --- |
|  | (3) |

Here, is the adversarially modified data, is the gradient of the MSE loss function, is the historical mean of the trading decisions, and is a scaling parameter called attack rate (Goodfellow et al., 2015).

* 1. Results

**Ein Bild, das Text, Diagramm, Reihe, Schrift enthält.

Automatisch generierte Beschreibung**This Section investigates the effects of the black-box attack proposed in Section 3 to the grid-scale battery problem presented in Section 2. In particular, Section 4.1 shows the how the attacks affect the operational schedules and Section 4.2 shows the statistics of the changes in profits for different attack rates and the three different regression models. This Section only considers attacks on the full decision-making process as the individual components like the EPF and the DSM models are unknown to the attacker, i.e., the attacker cannot evaluate the effect of the attack on the EPF performance or the sensitivity of the DSM model to the electricity prices directly.

Figure 2 Changes of the scheduling decisions for attack rates , , in relation to the actual day-ahead electricity price on July 1st, 2019.

* + 1. Storage schedules

Figure 2shows the schedule of the grid-scale battery storage on July 1st, 2019, for different attack rates. The figure relates the actual day-ahead electricity price with the state of charge schedule of the battery storage. The blue line labeled TCN shows the battery operation schedule obtained with the forecasts from the TCN without any perturbations. The schedule clearly shows how the battery is charged in the early morning between midnight and 4 am, when the electricity prices are low, and discharged between 5 am and 7 am, when the electricity prices peak. A similar cycle then repeats later in the day. With increasing attack rates, the operation of the electricity storage changes to less favorable schedules. In particular, the recharging, which previously started around noon and continued during the low-price hours of the afternoon, now starts earlier when the electricity prices are still high. For instance, for an attack rate of , the battery is charged between 6 and 7 am when the electricity price is at its peak. The schedules shown in Figure 2 highlight how small perturbations to the input data can lead to significant and harmful changes in the scheduled operation of the grid-scale electricity storage.

* + 1. Profits

Ein Bild, das Text, Screenshot, Diagramm, Schrift enthält.

Automatisch generierte BeschreibungUltimately, the adversarial attack in this work aims to turn the profitable operation of the battery storage to lower profits or even losses. Figure 3 shows the mean and variance bands of the profits obtained via DSM of the battery storage. The profits are computed for results for each day in 2019 and attack rates between 0 and 1 to obtain representative statistics. The results show decreasing profits for all three forecasting models. However, the decline in profits obtained with LASSO regression is significantly slower compared to the two neural networks, where the profits decrease quickly and even lead to losses. The drop of in profits confirms the observation from Figure 2 that the proposed black-box attack leads to a deterioration of the decisions made by the optimizer in Problem (P). The LASSO regression is fitted using a regularization of the scale parameters of the linear regression. Thus, the LASSO prediction is dominated by its bias term, i.e., a constant that is not affected by the adversarial attack. Therefore, the LASSO regression model is more robust towards the attacks. Meanwhile, the neural networks include many scaling parameters making the impact of small perturbations more significant. Notably, the profits obtained with LASSO forecasts without any attacks are significantly lower compared to the two neural networks. Hence, a trade-off occurs between the accuracy of the electricity price forecast and the robustness towards adversarial attacks.

Figure 3 Profits obtained in DSM of the grid scale battery storage for the three forecasting models and attack rates between 0 and 1.

* 1. Conclusions

This work considers a case of an adversarial attacker aiming to attack the combined decision-making process of EPF and DSM. In the considered case, the attacker has no access to the individual models and cannot intervene with any intermediate state of the decision-making process. The main proposal of this work is an ASM that emulates the decision-making process by training a regression model on the inputs and outputs of the decision-making process; the residual load forecasts that form the basis of the price forecasts and the decisions made via the optimization problem.

The results of the grid-scale battery case study show that small perturbations based on targeted gradient information can lead to significant damage in DSM. Notably, the attacker can design and implement the black-box attack without accessing the company's internal network directly. Instead, the data pipelines to and from the company need to be monitored and only the input data must be modified. Thus, the proposed attack poses a potential threat to the profitability of companies practicing DSM and additional work is required to develop defensive strategies.

* 1. Acknowledgments

The author gratefully acknowledges the financial support of the Kopernikus project SynErgie 3 by the Federal Ministry of Education and Research (BMBF) and the project supervision by the project management organization Projektträger Jülich.

References

Ave, G. D., Harjunkoski, I., & Engell, S. (2018). Industrial Demand Side Management Formulation for Simultaneous Electricity Load Commitment and Future Load Prediction. In *Computer Aided Chemical Engineering* (Vol. 44, pp. 1237–1242). Elsevier B.V. https://doi.org/10.1016/B978-0-444-64241-7.50201-9

Baader, F. J., Mork, M., Xhonneux, A., Müller, D., Bardow, A., & Dahmen, M. (2020). Mixed-Integer Dynamic Scheduling Optimization for Demand Side Management. In *Computer Aided Chemical Engineering* (Vol. 48, pp. 1405–1410). Elsevier B.V. https://doi.org/10.1016/B978-0-12-823377-1.50235-4

Cramer, E., Paeleke, L., Mitsos, A., & Dahmen, M. (2022). Normalizing flow-based day-ahead wind power scenario generation for profitable and reliable delivery commitments by wind farm operators. *Computers and Chemical Engineering*, *166*. https://doi.org/10.1016/j.compchemeng.2022.107923

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–11.

Morales, J. M., Conejo, A. J., Madsen, H., Pinson, P., Zugno, M., Grigoroudis, E., & Siskos, Y. (2014). Integrating renewables in electricity markets - Operational problems. In *Springer* (Vol. 205). https://doi.org/10.1007/978-1-4614-9411-9

Trebbien, J., Gorjão, L. R., Praktiknjo, A., Schäfer, B., & Witthaut, D. (2023). Understanding electricity prices beyond the merit order principle using explainable AI. *Energy and AI*, *13*, 100250. https://doi.org/10.1016/j.egyai.2023.100250

Xu, H., Ma, Y., Liu, H. C., Deb, D., Liu, H., Tang, J. L., & Jain, A. K. (2020). Adversarial Attacks and Defenses in Images, Graphs and Text: A Review. *International Journal of Automation and Computing*, *17*(2), 151–178. https://doi.org/10.1007/s11633-019-1211-x