Energy Informatics

RESEARCH Open Access

Towards reinforcement learning for vulnerability analysis in power-economic systems

Thomas Wolgast[1,2*], Eric MSP Veith[2] and Astrid Nieße[1,2]

From The 10th DACH+ Conference on Energy Informatics
Virtual. 13-17 September 2021

*Correspondence:
thomas.wolgast@uni-oldenburg.de
[1] University of Oldenburg, Ammerländer Heerstraße 114-118, Oldenburg, Germany
[2] OFFIS – Institute for Information Technology, Escherweg 2, Oldenburg, Germany

Abstract

Future smart grids can and will be subject of systematic attacks that can result in monetary costs and reduced system stability. These attacks are not necessarily malicious, but can be economically motivated as well. Emerging flexibility markets are of interest here, because they can incite attacks if market design is flawed. The dimension and danger potential of such strategies is still unknown. Automatic analysis tools are required to systematically search for unknown strategies and their respective countermeasures. We propose deep reinforcement learning to learn attack strategies autonomously to identify underlying systemic vulnerabilities this way. As a proof-of-concept, we apply our approach to a reactive power market setting in a distribution grid. In the case study, the attacker learned to exploit the reactive power market by using controllable loads. That was done by systematically inducing constraint violations into the system and then providing countermeasures on the flexibility market to generate profit, thus finding a hitherto unknown attack strategy. As a weak-point, we identified the optimal power flow that was used for market clearing. Our general approach is applicable to detect unknown attack vectors, to analyze a specific power system regarding vulnerabilities, and to systematically evaluate potential countermeasures.

Keywords: Ancillary service markets, Systemic weaknesses, Reinforcement learning, Economic attack, Artificial intelligence, Machine learning

Introduction

In recent years, attacks on the electrical power system have received more and more attention in scientific literature. Publications focus mainly on cyber-attacks and physical manipulation as attack vectors with the malicious intention of physical damage in the power system, e.g. Paul and Ni (2017); Ju and Lin (2018); Yan et al. (2017). However, apart from these malicious intentions, monetary incentives can be an even more important motivation to attack power systems. The so called increase-decrease

Springer Open

(inc-dec) gaming (Hirth et al. 2018) is one example for that. In inc-dec gaming, power plant operators exploit an expected congestion in the power system by strategically bidding on the energy market in a way to aggravate that congestion. Then, they sell the unused power plant's flexibility to the grid operator as profitable ancillary service to fix the previously aggravated congestion. Here, strategic bidding on the energy market is used as attack vector to increase profit on flexibility markets. Although not explicitly considered as attack strategy in literature, inc-dec gaming is an example that markets can be a gateway or incentive for manipulation. The example of inc-dec gaming also demonstrates a general issue of flexibility markets: Flexibility providers are possibly able to manipulate the physical power system in a way to increase the worth of their flexibility provision (Buchholz et al. 2021). This field of tension between the physical power system and coupled market systems, which we call a power-economic system, requires more research attention, thus motivating our research.

The first challenge is to draw a clear line between normal market behavior and an attack strategy. For this work, we characterize an economically motivated attack—or economic attack for short—as follows:

- Profit increase: It increases the expected profit of the market participant (otherwise not economically motivated).
- Health decrease: It decreases grid health, for example by inducing constraint violations (otherwise no attack).
- Repetitive effect: It is reproducible and done on a regular basis (otherwise no purposeful strategy).

It is important to note that different power-economic systems can have different vulnerabilities and therefore attack strategies that exploit them, considering that there are differences in the physical system, the operation mode, market integration, and so on. If similar strategies to inc-dec gaming can be applied to other settings, it would be a serious problem regarding grid stability and monetary costs for grid operators. If economic attacks are rational for a potential attacker, they will happen. Due to the current trend towards optimization and automation, intelligent agents will detect and exploit potential vulnerabilities, if they are economically profitable. No malicious intentions are required for that and possibly it will happen even without knowledge of a human operator. If that seems too far-fetched, imagine the scenario of a large virtual power plant with thousands of actuators that participates in multiple markets and is controlled by a machine learning algorithm trained on data amounts that are to big and complex to be understandable for the human operators. However, the dimensions and possibilities of such attacks are still mostly unknown. In this context, the general question is what kind of exploits and attack strategies are possible in power-economic systems. From the perspective of a single grid operator the question can be formulated more specifically: How can be found out if there is a strategy to exploit a specific power-economic system? That knowledge is required to evaluate countermeasures to protect the power system against attacks and their consequences.

To systematically search for unknown attack strategies and vulnerabilities, a methodological approach is needed that fulfills the following requirements:

- Unbiased: To find unknown attack vectors, as little knowledge and assumptions as possible about potential attack strategies are required.

- Explorative: Multiple attack strategies could exist for a single system. It is important to not only find the optimal strategy but all possible strategies, especially to evaluate countermeasures.
- Sequential: Multiple actions could be necessary in a specific sequence to realize an economic attack, see for example (Ju and Lin 2018). A systematic search needs to consider long sequences of actions as attacks.
- Complexity: The methodology needs to deal with complex power systems, markets, and potentially information and communication technology (ICT), which are interconnected and can all serve as potential attack vector.

Such a methodology would not only be applicable to find new attack strategies but would also enable deeper investigation of already known strategies like inc-dec gaming. Our contributions are as follows: In the following related work section, we identify deep reinforcement learning (DRL) as a suitable empirical approach to search for attack strategies in power systems that fulfills the discussed requirements. By the example of controllable loads in a reactive power market setting, we demonstrate how DRL can be used by grid operators as a methodology to identify systemic vulnerabilities as a proof-of-concept. Further, we use the approach to identify a new attack strategy that utilizes controllable loads to exploit a reactive power market by attacking it[1]. In this specific work, we focus on interdependencies between physical power systems and flexibility markets. However, the general approach is broadly applicable to vulnerability analysis in power systems.

Related work

The cyber attacks on the Ukrainian power system in December 2015 (Case 2016; Styczynski and Beach-Westmoreland 2016)—with further attacks until and beyond 2017 (Reuters 2017)—have demonstrated that energy systems are both valuable and vulnerable targets. In general, cyber attacks are the most researched kind of attacks on power systems. An important class of cyber attacks is false data injection to manipulate state estimators (Liang et al. 2017). Xie et al. 2010 demonstrate how false data injections can be used to manipulate energy markets and to increase market profit this way. Chen et al. 2019 demonstrate how load forecasting can be manipulated with minimal system knowledge using machine learning and gradient estimation methods.

Another class of attacks are physical manipulations of the system. Ju and Lin 2018 investigate the damage potential of compromised assets that actively manipulate distributed Volt/VAR control. Ni et al. 2018 and Yan et al. 2017 utilize reinforcement learning (RL) to identify critical sequences of topology attacks.

The example of inc-dec gaming demonstrates how markets can be used for manipulation and to generate additional profit. Inc-dec gaming is investigated especially by Hirth and Schlecht for the German power market (Hirth et al. 2018), but similar analysis for the intraday market of Great Britain was done by Konstantinidis and Strbac 2015. Altogether, comparably few publications consider attacks for economic reasons.

Attacks with distinguishable actors and known attack motivation can often be modelled as games. Farray et al. 2016 consider switching attacks, whereas Wei et al. 2018 implement a defense mechanism using Nash equilibria. While game theory can be applied to formalize actor behavior and identify equilibria in well-known systems, their application

[1] All the source code can be found at: https://gitlab.com/digitalized-energy-systems/scenarios/reactive_market_attack

is limited in systems with unknown actors and complex interdependencies of the subsystems. For example, this was shown for investment games by Kiedanski et al. 2019: They demonstrated that by adding more details to the physical model behind the game theory formalization, the results changed significantly. This shows that with a purely game-theoretic modeling, interactions in complex power-economic systems cannot be evaluated sufficiently. To this end, considering market attacks, Spooner and Savani 2020 employ a game-theoretic adaptation of a standard mathematical model of market making, but train RL agents as robust actors in the face of market attacks, thus combining RL generated models and game theoretical market modeling.

Multiple of the mentioned publications use RL-based approaches to investigate attack strategies. Generally, RL has seen a renaissance through the hallmark publication by Mnih et al. 2013, which resulted in the field of DRL. DRL has since then firmly established itself in dealing with extremely complex tasks, culminating in the enormous successes of AlphaGo (Silver et al. 2016), AlphaZero (Silver et al. 2017), and MuZero (Schrittwieser et al. 2019).

RL is the general paradigm of optimizing the policy of an agent to interact with its environment to maximize some reward. RL has a very general interface to the environment. It requires only a minimum set of assumptions about the expected agent behavior. Applied to the power system, only reward, sensors and actuators need to be defined, if a simulator of the power system is given as environment. This enables a relatively unbiased search for attack strategies. Further, DRL often deals with large action and state spaces while optimal actions are by definition unknown. That is why extensive research is done regarding exploration of the state and action space. That allows to find multiple different attack strategies, if more than one exists. The example of beating the game of Go demonstrates how DRL can deal with long-term sequential decision making in extremely complex environments (Silver et al. 2016; Silver et al. 2017; Schrittwieser et al. 2019). Therefore, DRL fulfills all the necessary requirements for an unbiased search for attack strategies in very complex power systems, as discussed in the "Introduction" section.

However, it was shown that DRL-based methods are not used to analyze power-economic systems in order to identify unknown attack strategies that exploit power grid control or energy markets (Veith et al. 2019). This is in stark contrast to what has been discussed and referenced above. Up to now, the scope has been limited to optimize or analyze *known* or *foreseeable* attack strategies.

Our approach is to explicitly let agents derive their own attacks, also limiting their knowledge of the power grid to their local view. To this end, we propose a DRL-based analysis in the sense of adversarial resilience learning (ARL) (Fischer et al. 2018; Veith et al. 2020), where an attacker agent with minimal domain knowledge learns to exploit the system and to identify vulnerabilities this way.

The power-economic system under attack

To investigate in an exemplary case study how DRL can be used to find attack strategies and therefore exploitable weaknesses in power-economic systems, we model a medium voltage (MV) distribution grid where the distribution system operator (DSO) procures its reactive power demand from a local reactive power market. Reactive power can be used by the grid operator to ensure constraints satisfaction–e.g. the voltage band–or to optimize the system state. From a RL perspective, that power-economic system is the environment of the learning agent, including the physical power system, the local market, distributed energy resourcess (DERs), and non-attacker market participants. We use local reactive power markets as example here because the local character of reactive power allows for small models with clear system boundaries. Further, lots of research is done how to design future reactive power markets (Jay and Swarup 2021), but if these markets are attackable was not investigated yet. Our hypothesis is that the tight coupling with the physical power system enables potential attackers to manipulate the system in a way to increase profits on the market. We test that hypothesis by providing controllable loads and generators to some attacker agent. While the loads can be used to manipulate the system, the generators' profit on the reactive power market serves as incentive.

In the following sections, we discuss the modelling and our assumptions regarding the physical power system, the reactive power market, the non-attacker units and agents, and also the controllable loads of the attacker agent.

Physical power system model

To model the power system, we use the python package *pandapower*[2] that provides easy to handle power system models and an optimal power flow (OPF) algorithm that we use for the reactive power market (Thurner et al. 2018). Pandapower also provides an interface to the *simbench*[3] package that provides some modern benchmark networks with one-year time-series data in quarter-hourly resolution for loads and generation. We use them in the back-end of the environment to generate realistic system states.

Generators and loads

For the reactive power capability, we assume all generators to be designed with a $\cos(\varphi)$ below one, resulting in a maximum apparent power of

$$s_{\max} = \frac{p_{\max}}{\cos(\varphi)}, \tag{1}$$

with maximum active power feed-in of $p_{\max}$. Further, we assume that they can set their reactive power feed-in independently from active power. The reactive power capability $q_{\max,t}$ is computed at each time step t by

$$q_{\max,t} = -q_{\min,t} = \sqrt{s_{\max}^2 - p_t^2}. \tag{2}$$

While the active power feed-in p_t of all generators is taken from the simbench time-series data, reactive power is set solely by the market rules of the reactive power market, which we discuss in the "Optimal power flow and reactive power market" section. The non-attacker loads simply adopt the active and reactive power values from the time-series. The controllable attacker loads are discussed in the next section.

Controllable attacker loads

Lots of load types are controllable to some extent, for example cooling units, heat pumps, or electric vehicle charging, but that flexibility is barely offered on flexibility markets, especially from smaller units in distribution systems (Minniti et al. 2018). Further, loads in distribution systems are only barely monitored or even remote controllable, which makes

[2] Pandapower documentation: https://pandapower.readthedocs.io
[3] Simbench documentation: https://simbench.readthedocs.io

them almost a blackbox for the grid operator. Altogether, controllable loads are especially suited to perform systematic attacks on distribution systems.

To allow for controllability, we model the attacker loads as energy storage systems. Their maximum power demands $p_{\max}$ and $q_{\max}$ are chosen as the maximum value in the simbench time-series. Their respective minimum power is zero, i.e. no power feed-in is possible, in contrast to normal storage systems. At each time step t, the load l has a storage loss $p_{\text{loss},l,t}$ that is equivalent to the original active power demand of the time-series. This way, a timely shift of the power demand is possible, but it is ensured that the average demand remains unchanged. Further, the storage loss is not constant, which emulates fluctuating user behavior. Note that we use *storage* and *attacker load* interchangeably in the following. We define the storage capacity C of the attacker load l as:

$$C_l = \tau_{\text{shift}} \cdot p_{\max,l} \tag{3}$$

The value of τ_{shift} describes how shiftable the load is and is heavily technology dependent (Gils 2014). The attackers can control the power demand relative to its maximum value by a scaling factor a. Thereby, we assume that active and reactive power are coupled and not independently controllable:

$$p_{l,t} = a_{l,t} \cdot p_{\max,l} \tag{4}$$

$$q_{l,t} = a_{l,t} \cdot q_{\max,l} \tag{5}$$

At every step t, the storage level F is increased by active power demand p_t and decreased by the underlying simbench load profile $p_{\text{loss},l,t}$. By deviating from the simbench load profile, the storage can be filled or emptied.

$$F_{l,t+1} = F_{l,t} + \frac{p_{l,t} - p_{\text{loss},l,t}}{C_l} \tag{6}$$

To prevent the storage level from leaving the permitted range, the load demand $p_{l,t}$ gets increased to prevent underflow or decreased to prevent overflow, if required, at each step respectively. This clipping also ensures that the average and total load demand remains unchanged.

Optimal power flow and reactive power market

The reactive power market model is OPF-based, following the standard literature on reactive power market. Noteworthy are for example (Zhong and Bhattacharya 2002 and Amjady et al. 2010). The most important characteristic of reactive power markets is that reactive power must be provided locally close to the location of demand, which makes its worth vary from bus to bus. Therefore, it is sub-optimal for the DSO to simply accept the cheapest reactive power offers. The common method – as in the mentioned publications – is to perform an OPF calculation for market clearing to determine the cost-optimal reactive power provision for the DSO that satisfies all system constraints. The OPF usage for market clearing reflects the mentioned tight coupling between power system and reactive power market. For this reason, the OPF and the market are handled together in this section.

As system constraints, we consider maximum and minimum voltage levels and maximum line and trafo loadings. We assume fixed active power feed-in of the generators from some previously cleared energy market. That market does not need to be explicitly modelled here, which results in a pure ORPF. Further, we assume that the DSO has no own actuators like capacitor banks and is therefore fully dependent on market-based reactive power provision from generators. We are aware that not all potential constraint violations can be cleared with reactive power provision only. However, this approach was chosen to decouple it from the normal energy market and to have clear system boundaries this way. If the OPF fails for the mentioned reasons, the system constraints gradually get relaxed until the OPF calculation succeeds. That allows us to determine if manipulations of the system result in a higher number of constraint violation in comparison to a non-attack scenario. That would not be possible, if the OPF always succeeded.

Regarding the reactive power market, we make the following assumptions:

- The market is pay-as-bid.
- All generators participate in the reactive power market with their full capability $q_{\mathrm{min/max}}$ (see Eq. 2).
- Market clearing is performed at each step. That is a shorter time-scale than would be realistic, but simplifies OPF calculation drastically, decouples the steps from each other, and ensures optimal set-points at every step.
- After market clearing, the generators are obliged to provide the requested reactive power that results from the OPF.

Simulation procedure

In the previous sections, the power-economic system was defined, which is equivalent to the RL environment. Since some updates within the environment depend on each other, the following listing defines the exact order of the updates that happen in a single step t.

1 Agent gives actions to environment.
2 Environment is set to the next step ($t = t + 1$).
3 Non-attacker units' active power is set to next step .
4 Attacker-loads are set and their internal storage levels updated.
5 Constraints for current step are calculated (see Eq. 2).
6 OPF is performed to determine optimal reactive set-points.
7 If OPF failed: Relax constraints and jump back to step (6).
8 Generators get paid for their service and reward is calculated.
9 Agent receives observation and reward.

The reinforcement learning problem

In this section, we define the RL problem by choosing the reward function, the action space, and the observation space. Further, we discuss the train-test split of the simbench data set and present the utilized DRL algorithm.

Observation space

We assume a single attacker that is in control of a sub-set of nodes N^* in the grid and controls all generators G^* and loads L^* that are connected to these buses. The superscript asterisk always indicates the attacker here. For the observation space, we generally presume knowledge of the local variables for own buses of the attacker, which makes the environment partially observable. For every time step t, the attacker agent gets the following information as observation:

The agent observes all local voltage magnitudes u_n of the controlled nodes N^*,

$$O_1 = u_{n,t-1} \; \forall \; n \in N^* \tag{7}$$

current active and reactive power set-points of all controlled loads and generators,

$$O_2 = (p_{l,t-1}\,,\; q_{l,t-1}) \; \forall \; l \in L^* \tag{8}$$

$$O_3 = (p_{g,t-1}\,,\; q_{g,t-1}) \; \forall \; g \in G^* \tag{9}$$

and the storage levels of all attacker loads.

$$O_4 = F_{l,t-1} \; \forall \; l \in L^* \tag{10}$$

Further the agent receives the information of the next active power values of the load profile, which is equivalent to the storage loss of the current step. These are the values that the agent would set, if it would act normally.

$$O_5 = p_{\mathrm{loss},l,t} \; \forall \; l \in L^* \tag{11}$$

Further, the next active power values of the generators are known, which are required to estimate their possible reactive power flexibility on the market.

$$O_6 = p_{g,t} \; \forall \; g \in G^* \tag{12}$$

Finally, the agent has information about the current time in the year, week, and day. To consider the cyclical characteristics, we encode them as sine/cosine pairs respectively, which makes six additional observations. The yearly features distinguish seasons, weekly features workdays and weekends, and the daily features encode the day-night cycle. Sine and cosine are both required to make the features unambiguous.

$$O_7 = \sin(2\pi \cdot \frac{t \,\%\, tf}{tf}) \; \forall \; tf \in TF \tag{13}$$

$$O_8 = \cos(2\pi \cdot \frac{t \,\%\, tf}{tf}) \; \forall \; tf \in TF \tag{14}$$

With % as modulo function and the three time-frames

$$TF = \{4 \cdot 24,\; 4 \cdot 24 \cdot 7,\; 4 \cdot 24 \cdot 366\}. \tag{15}$$

In addition, we provide the agent with the observations from the previous two steps, which is a usual procedure in DRL to make system dynamics observable, e.g. in Mnih et al. (2013). Overall, the total number of observations is:

$$n_{\mathrm{obs}} = 3 \cdot |O| = 3 \cdot (|N^*| + 3|G^*| + 4|L^*| + 6) \tag{16}$$

Action space

As action space A, the attacker can only control the ratio $a_{l,t+1}$ of the load power in the continuous range [0, 1] relative to the maximum possible load power of every load l as seen in Eq. 4.

$$A = a_{l,t+1} \in [\,0,1\,] \; \forall \; l \in L^* \tag{17}$$

$$n_{\mathrm{act}} = |L^*| \tag{18}$$

The attacker generators explicitly have no possibility to manipulate the environment in any way, as discussed earlier.

Reward function

The total market profit ϕ is simply the summed up profit of all attacker generators G^* and can be directly used in the reward function as an incentive for the agent.

$$\phi = \sum_{g \in G^*} \phi_g \tag{19}$$

The active power costs of the loads do not need to be considered, because on average they remain unchanged compared to the underlying load profile. A capacity price for maximum yearly power demand does not need to be considered either, because the maximum active power demand stays unchanged compared to the original load profile as well. However, these assumptions are only valid because invalid actions get clipped by the environment automatically. To consider this and to give feedback about invalid actions to the agent, an additional penalty ψ function is introduced. The penalty increases quadratically with the difference between clipped actions a_l and the intended actions a'_l.

$$\psi = 3 \cdot \sum_{l \in L^*} (a_l - a'_l)^2 \tag{20}$$

The penalty factor three was chosen in a way that the penalty is in a similar range as the profit so that it serves as learning incentive for the agent, but not to big to dominate the learning process. Our experiments showed that the penalty increases learning success significantly. Overall, the reward of step t is the market profit minus the penalty:

$$r_t = \phi_t - \psi_t \tag{21}$$

Train/test-split

The training steps are drawn from the simbench time-series data with a length of 35,136 time steps. Consequently, at some point in training, the steps will repeat, which bears the risk that the attacker over-fits to the time-series data. To validate the results, a part of the profile data is taken as test data that is not used in training. Overall, we chose about 23% of the data in the form of twelve one-week episodes at the beginning of each month and therefore equally distributed over the year, which makes 8028 test steps in total. That was done to ensure that all weekdays and seasons are equally represented in the data.

The TD3 algorithm

We chose Twin-Delayed DDPG (TD3) (Fujimoto et al. 2018) as DRL algorithm for our agent. TD3 itself is based on Deep Deterministic Policy Gradient (DDPG), which makes it suitable for the continuous action space at hand. As an extension to DDPG, TD3 addresses the over-estimation of state values $V(s)$ with a second Q-network, an idea well-known from Deep Q-Networks (DQN) (van Hasselt et al. 2016). Further, TD3 adds delayed policy updates and noise to the target action. Hyperparameters and information about random seeds can be found in Appendix 2.

Case study

For our proof-of-concept case study, we will define a scenario and some evaluation metrics to test our approach. Afterwards, we apply the TD3 algorithm to the scenario and present our results briefly.

Scenario

As exemplary power system for our case study, we investigate the 97-bus rural MV simbench system that is shown in Fig. 1. We assume that the attacker controls the eight adjacent nodes 61 to 68 at the end of feeder 6 in the west and that it controls all nine loads and nine generators at these nodes. Overall, the attacker controls about 11.8% of installed load capacity and 4% of installed generation capacity. That makes 231 observations and 9 actions in total, according to Eqs. 16 and 18. We intentionally consider a larger power system with comparably few attacker units to investigate if the market can be exploited by attackers that can only locally observe and influence a minor part of the system, which is more realistic than the presumption of an all-knowing and globally acting attacker agent. Further, we assume all loads to be up-scaled by factor three. Although quite unrealistic, we ensure this way that the attacker units are actually big enough to manipulate the system so that constraint violations can emerge. That is done for all loads so that the attacker loads do not have an advantage in comparison. We assume $\cos(\varphi)$ of all generators to be 0.95. For the shiftable attacker loads, we exemplarily assume τ_{shift} as a time-frame of 1.5 h. That is small for an electric storage water heater, but large for a refrigerator (Gils 2014). For the reactive power market, we assume all generators to have a quadratic cost function of 250 €/mvar^2/h to consider internal losses, based on (Samimi et al. 2015). Further, all generators have a fixed profit margin of 100% and therefore expect 500 €/mvar^2/h as payment. That includes attacker generators, because this work completely focuses on physical attacks on the system, while the market behavior remains unchanged and the profit is only the incentive for attacks. For the OPF, we utilize the pandapower AC-OPF and define a voltage band of ±5% around reference voltage and maximum line and trafo loading of 60% as system constraints. In case of failure, we gradually relax the constraints by ±0.5% for the voltage band and 5% for line and trafo loading as described in the "Optimal power flow and reactive power market" section.

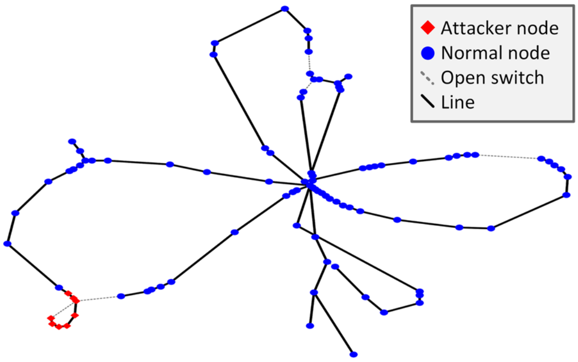

Fig. 1 The attacked 97-bus rural MV simbench system including the attacker nodes

Evaluation metrics

We introduce five metrics to evaluate if the agent learned some economically successful strategies and if these strategies can be considered as attacks, following our characterisation in the beginning. The summed market profit Φ of the attacker generators G^* over all test time steps T is used to measure economic success.

$$\Phi_{G^*} = \frac{1}{4} \sum_{t \in T} \phi_{G^*,t} \tag{22}$$

The factor 1/4 is necessary because the simulation is stepped in quarter hour steps. Note that by Eq. 19, the profit depends only on the reactive power feed-in q_g, which is set by the grid operator and cannot be influenced directly by the attacker agent. Second, the attacker market share σ_{G^*} measures the competitiveness of the attacker generators G^* relative to all generators G.

$$\sigma_{G^*} = \Phi_{G^*} / \Phi_G \tag{23}$$

Normal market outcome on an ancillary service market is expected to provide flexibility to the grid operator and should therefore improve grid health, e.g. by preventing constraint violations. To distinguish between economic attacks that are harmful to the power system and expected normal market behavior that only maximizes profit without harming the system, we measure the number of system constraint violations. That is done over all nodes N in the system and all branches B. Constraint violations are categorized by undervoltage (index uv), overvoltage (ov), and branch overload (ol) situations.

$$n_{\mathrm{uv}} = \sum_{t \in T} \sum_{n \in N} u_{n,t} < 0.95\,\mathrm{pu} \tag{24}$$

$$n_{\mathrm{ov}} = \sum_{t \in T} \sum_{n \in N} u_{n,t} > 1.05\,\mathrm{pu} \tag{25}$$

$$n_{\mathrm{ol}} = \sum_{t \in T} \sum_{b \in B} s_{b,t} > 0.6 \cdot s_{b,t}^{max} \tag{26}$$

The choice of system constraints was discussed in the previous section. These metrics alone are not sufficient to determine if successful attacks could be learned, because even if the attacker acted normally, some profit and constraint violations could emerge. Therefore, we compute all metrics in comparison to a baseline scenario. The baseline scenario is that all attacker loads act normally in the way that they follow the underlying load profile exactly as the non-attacker loads do. The difference between the metrics for the attack-scenario and the baseline scenario shows the impact of the agent actions. We trained 15 TD3-agents with different random seeds independently from each other for 200 tsd. steps with randomly selected four-day episodes from the training data alone. We repeated the same procedure with an Actor Critic using Kronecker-Factored Trust Region (ACKTR) algorithm (Wu et al. 2017), but did not find any results that deviated significantly from the baseline, which is why only the TD3 results are shown in the following. The learning curve of the TD3 experiments is shown in Fig. 2. The plot shows an overall monotonously increasing average return per episode.

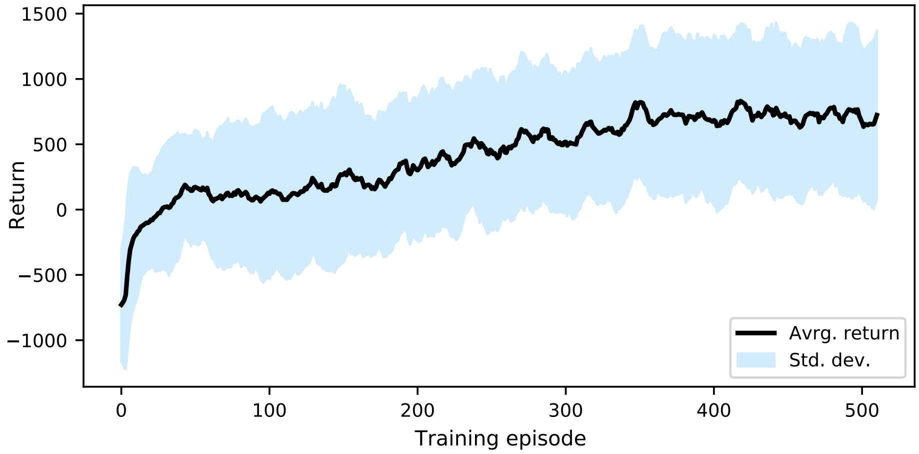

Fig. 2 Learning curve of the 15 agents over 200 tsd. training steps with average return and standard deviation averaged over 10 episodes

Results

After training, the agents were tested on the 12-week test data that was not seen in training. The evaluation metrics from the test runs are listed in Table 1 in comparison to the baseline scenario and also visualized in Fig. 3.

Table 1 shows that the average attacker profit increases drastically by about 323.2% on average and the market share increased by about 39.5% compared to the baseline. The relative additional profits are more relevant here than the absolute values, because these depend heavily on the chosen parameters like the profit margin. In the baseline scenario almost no constraint violations happened. Only three times a line was overloaded, which can be traced back to the up-scaling of the loads. In comparison, the learned strategies resulted in 2216.2 overloaded lines, 3103.27 undervoltage situations, and even 2.07 overvoltage situations on average. It must be remembered here that only nine loads were allowed to change their behavior strategically and that all other loads and all generators still maintained their baseline behavior. That makes these nine loads solely responsible for the violations. Fig. 3 visualizes the metrics' distribution. Overvoltage situations were omitted, because they barely ever happen. The plot together with Table 1 shows that all 15 agents achieved drastically higher profit than the baseline without exceptions. The same is true for the other metrics. Further, the plot indicates a high correlation between profit and the other metrics. Thereby, the correlation with the constraint violations is far stronger than with the market share. Both Fig. 3 and Table 1 show a high variance over the experiments for all metrics.

Table 1 Averaged evaluation metrics of the 15 trained agents, including standard deviation, in comparison with the deterministic baseline applied to the 8028 test steps

Metric	Baseline	Learned
Profit Φ_{G*} [€]	6,529	27,631 ± 3,064
Market share σ_{G*} [%]	6.79	9.47 ± 0.43
Overload n_{ol}	3	2216.2 ±339.85
Undervoltage n_{uv}	0	3103.27 ± 467.27
Overvoltage n_{ov}	0	2.07 ± 1.81

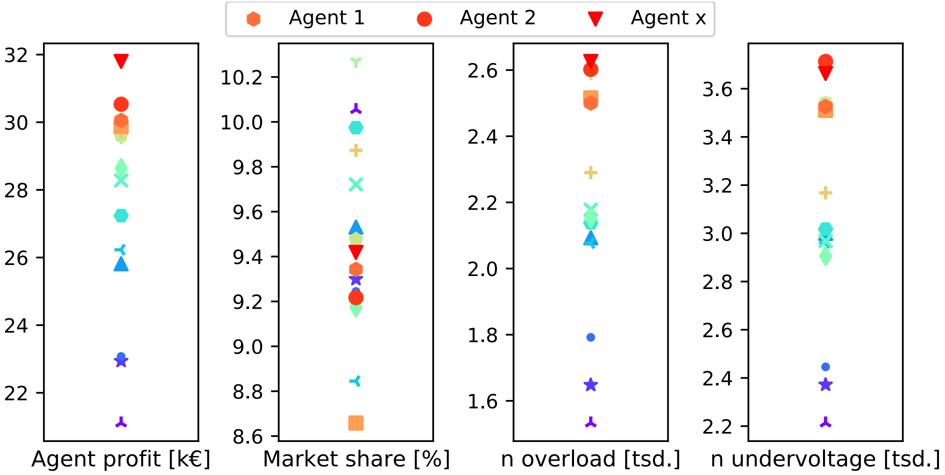

Fig. 3 The four metrics profit, market share, and number of overload and undervoltage situations for all experiments. The colors are sorted by profit. Reddish colours indicate higher profit, blueish colors indicate lower profit. For comparison with the baseline, refer to Table 1

To investigate how the agents achieved these profits and violations, we show a three day window of the may test episode for the agent with the maximum profit of about 31.7 k€ in Fig. 4. It visualizes the agent's profit and its actions compared to the baseline. Further, it shows the course of the minimum voltage in the attacker area in subplot c)

$$f(x) = \min(u_{n,t} \ \forall \ n \in N^*) \tag{27}$$

and also the maximum line load in the whole feeder B^* that the attacker units are connected to in subplot d):

$$f(x) = \max(\frac{s_{b,t}}{s_{b,t}^{\max}} \ \forall \ b \in B^*) \tag{28}$$

The min/max values are visualized to make extreme cases and possible constraint violations visible. This specific agent and time-frame was chosen quite arbitrarily for illustration, because it shows some noticeable traits. However, the general results and learned strategies are the same for all agents and test episodes over the year.

The summed profit of the attacker generators in subplot a) shows multiple spikes during daytime of all three days. These spikes are dramatically higher than the baseline profit, which is almost constant, but shows some slight increase at daytime as well. At the second day, no additional profit could be generated in the evening hours, in contrast to the other two days.

Subplot b) shows the summed load ratio of the nine loads that were given to the learning agent, scaled to the range [0%, 100%]. The baseline load sum shows a characteristic day/night cycle with high load at day and lower load at night. Further, the baseline load sum does not take extreme values well below 20% or above 60%. The attacker loads follow the baseline course closely on average, but oscillate drastically around it, covering the full range from 0% to 100%, and resulting in drastic spikes in both directions. These spikes are correlated to the ones in profit and occur mainly at daytime. However, the action spikes are not perfectly correlated with the profit spikes, resulting sometimes in smaller profit spikes, e.g. at around 12:00 on the second day, or none at all, as can be seen in the evening of the second day or around 20:00 on the third day.

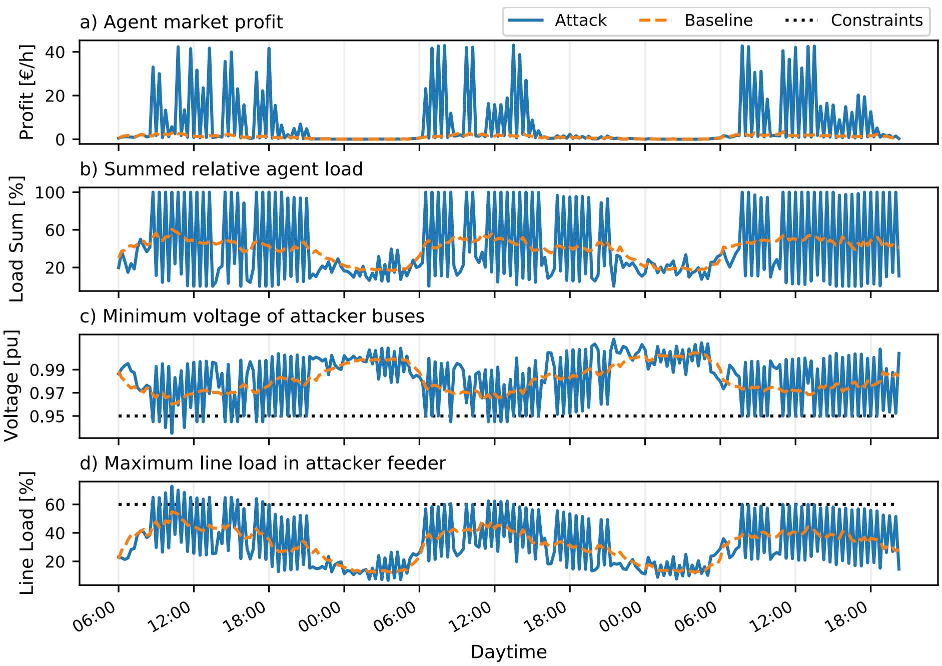

Fig. 4 Three days of the test data comparing the learned and the baseline scenario regarding a) agent profit, b) summed agent load, c) minimum voltage in the attacker area, and d) maximum line load in attacker feeder

The minimum voltage magnitude of the attacker buses is shown in subplot c). Again, we see a day/night cycle in the baseline that is negatively correlated with the load sum this time, resulting in lower voltage at daytime and higher voltage at night. Similar to the load sum, the voltage values of the attack scenario oscillate around the baseline values. Often, that results in voltage values that exactly match the voltage constraint of 0.95 pu or even violate it. The voltage values below the threshold imply that the OPF was not successful and that the system constraints had to be relaxed. All these observations are valid for the maximum line load as well, which is shown in subplot d), except that the line load is positively correlated with the load sum and the profit again. Further, it can be noticed that the load spikes not always match with voltages or line loads that intersect their respective constraint, for example after 15:00 on the second day or 20:00 on the third day. Exactly in these cases there is no profit spike correlated with the load spike. However, in reverse, constraint violations not always correlate with extreme profit spikes. For example at around 10:00 on the first day, both constraints are heavily violated, but the profit spikes are comparably small.

Discussion

In the following, we discuss the presented results. First, we discuss the specific strategy that was found by the DRL agents. Further, we evaluate our approach on a general level regarding its success in the case study and shortcomings.

The attack strategy

Figure 2 clearly shows that the agents learned a policy that improves their returns and that the observed behavior is not a result of chance. The metrics in Table 1 and Fig. 3

demonstrate that the learned strategy increases market profit drastically compared to the baseline. Further, the market share increased, which means that the agents were able to gain a competitive advantage over other reactive power providers. The same strategy also results in numerous constraint violations within the power system. Since the agents were tested on data that they have never seen before, it can be concluded that the learned strategy is applicable to unknown situations by utilizing the local observations, in contrast to just remembering the best actions from experience. In summary, the learned load shifting strategy is reproducible, profitable, and impacts grid health negatively. Altogether, that classifies it as economic attack as it was characterized in the "Introduction" section. However, the extreme extent of the results may seem surprising, since the attacker loads do not participate in the market, while the attacker generators participate in the market but cannot change their behavior in any way.

To understand how the attacker increased its profit that drastically, the short test-frame in Fig. 4 gives a good intuition about the attack strategy and why it works. The clear correlations between load sum and constraint violations show that the attacker is able to manipulate the power system so that constraint violations would occur. However, the OPF tries to prevent these violations and buys reactive power on the market, because that is the only flexibility option it has. Since the constraints are modelled as hard constraints that must be adhered to—which is common for the OPF—the grid operator is even forced to do so. This results in voltages and line loads that exactly match their respective constraints. In consequence, the attacker can set all its actuators to 100%, bring the system close to failure, and force the grid operator to buy flexibility on the market. That increases profits of the attacker generators without them being directly involved in the attack. Therefore, the violations are not only correlated to the profit increase—which was already visible in Fig. 3—but are a causal reason for it.

The locality of reactive power gives local providers natural market power (Zhong and Bhattacharya 2002). Generators in the energy system are often considered to have market power, when they are *must-run*, which means that their power feed-in is essential for operation of the power system (de Souza et al. 2001). An important factor are the overloaded lines: While the voltage violations require reactive power feed-in as countermeasure, the overloaded lines prevent the OPF from using reactive power from other generators that are further away. In other words, the attacker manipulates the system such that its generators become *must-run*. That results in market power and explains the extreme magnitude of the profit spikes. Conversely, that also explains why some attacks fail and why these are correlated with non-violation of the lines, e.g. the afternoon of the second day.

The oscillating behavior of the controllable loads is a result of their modelling as storage systems. When all loads are set to full power, the internal storage level reaches its maximum fast. Because of that, the agent needs to reduce the load power below the baseline regularly to recover, so that further attacks are possible in the future. Daytime attacks are more profitable, because the attacker recovers faster after an attack and also because the higher demand of non-attacker loads during the day makes the system easier to attack.

What can we learn from this on a general level? First of all, OPF-based reactive power markets can be exploited by strategic induction of constraint violations into the system and then providing the solution for it, as we showed here by the example of controllable loads and generators in the distribution grid. We discussed the theoretical possibility of that vulnerability already in (Buchholz et al. 2021), but it was unclear if it could work and was shown in simulation for the first time here. The OPF proved to be the weak-point, because its predictability and adherence to the hard constraints could be exploited, although being cost-optimal in the mathematical sense.

The results reinforce what is discussed regarding inc-dec gaming: Flexibility markets can incite unwanted behavior, if the market design is flawed (Hirth and Schlecht 2019). Therefore, this problem is of general nature and not specific to reactive power markets or congestion management, as we discussed earlier here and in Buchholz et al. (2021).

Our scenario of up-scaled loads may seem unrealistic. However, we believe that different power systems, energy and flexibility markets, and ICT in diverse combinations and implementations under various different regulation systems worldwide breed the possibility of an infinite amount of manifestations of attack strategies as the one that was found here. Therefore, if such attacks are possible in theory, it can also be assumed that they are possible in some energy system in the world now or in the future. Automated analysis tools are required that are able to systematically search for successful attack strategies in specific power systems to reveal underlying systemic weaknesses this way. With our work, we provide a first step towards that long-term vision.

Vulnerability detection with DRL

Regarding our proposed methodology, we will discuss if DRL is actually suitable to find systemic weaknesses in power systems by attacking them with a learning agent. The simulation results demonstrate that this is possible. The learning success of multiple independently trained agents not only shows the existence of the system vulnerability, but also demonstrates that it is exploitable under the given assumptions and local observation. The second outcome of the training is the learned strategy itself, which gives clues about the underlying problem and therefore potential countermeasures. This work focuses on the attack and does not discuss countermeasures in detail, but from the results it can be presumed that the grid operator could remove the vulnerability with a less overloaded network and soft constraints in the OPF.

This work demonstrated the approach by using a minimal example with some very restricting assumptions in a setting that we expected to be attackable as a proof-of-concept. These assumptions were helpful to reduce computational effort and to keep complexity of the scenario manageable. However, that does not question the approach itself. The actual attack strategy was autonomously found by the RL algorithm and the first time shown in simulation here. Another advantage of DRL is the algorithms' ability to cope with complex observation/action spaces with non-linear interrelationships. To analyze a power system thoroughly and to find attack strategies that were not even suspected beforehand, drastically more complex experiments need to be conducted. Imagine for example an attacker that controls the generators' active power provision as well, or learned to bid on the market, or other combinations of attacker loads and generators. For all these possibilities, assumptions were made here to keep the example simple.

That directly leads to the disadvantages of DRL for vulnerability detection. The system complexity of power systems in combination with markets or even ICT may result in extreme computational effort for learning. A second drawback can be seen in Table 1 and Fig. 3. Although all TD3 agents increased profit compared to the baseline, the variance of the metrics over all experiments is very high, which results from high randomness in DRL experiments (Henderson et al. 2018). Therefore, our approach requires multiple experiments to generate meaningful results, which increases computational effort again. Further, we repeated the experiments with ACKTR agents. These did not achieve any profit gain compared to the baseline. Consequently, multiple learning algorithms should be tested. Otherwise, there is a risk of overlooking an attack vector out of pure chance because the wrong DRL algorithm or bad hyperparameters were chosen. The third big drawback of DRL is the lack of explainability. In this work, we visualized the resulting attack strategy in Fig. 4 and interpreted it afterwards. However, this is only possible for very simple scenarios. And still, the interpretation contained a certain amount of speculation. Further, it remains unclear what exact observations in what combination the agents used to derive their actions from. Altogether, approaches are required to ensure interpretability of DRL algorithms. That is the field of explainable artificial intelligence (XAI) or rather explainable reinforcement learning (XRL), which gains more and more relevance (Puiutta and Veith 2020).

Conclusion

In this study, we showed how DRL can be used to find vulnerabilities in a power system by searching for successful attack strategies. We demonstrated a proof-of-concept in the form of a simple power-economic system, namely a MV distribution grid in which the grid operator procures reactive power from a reactive power market. The agent learned to utilize controllable loads to attack the system, exploit the market this way, and increase its profit drastically; a strategy that was not shown in simulation before.

Besides the exemplary use case of this work to find a new attack strategy, we expect our methodology to be applicable to diverse other settings and scenarios in research and industry, because it allows to find vulnerabilities automatically with comparably few assumption about potential attack strategies. Possible applications are for example:

1 As a research tool to find and analyze new general classes of attack strategies in power systems, as we did in this work by the example of the economic attack on reactive power markets.

2 Grid and market operators can use the approach in practice to test their current or future systems regarding potential vulnerabilities.

3 If vulnerabilities are found, countermeasures and alternative settings can be evaluated systematically to search for the ones that are most robust. Examples are power system planning, integration of ICT, regulation design, robust market design, and evaluation of operation strategies.

These examples demonstrate how the high-level idea of learning attack strategies can be used in various settings that reach far beyond the scope of this paper. Research in all these directions is important to design future power systems as robust as possible. In this work, we emphasize especially the market perspective in research of power system attacks, because markets provide a dangerous economic incentive for them.

Outlook

Interpretability of agent behavior and data is necessary to search for more complex and large-scale strategies. Therefore, XRL and automatic data analysis tools like anomaly detection are important future research directions. Automatic data analysis to find attack strategies is only possible, if a clear line between attack and normal behavior can be drawn and measured with clearly defined metrics. As discussed earlier, it is difficult to draw such a clear line for economic attacks because attack and rational market behavior are equivalent in this case. In this study, we used the non-attacker load behavior as baseline for comparison, but such a possibility may not always exist. A discussion is required about what acceptable normal behavior and what unwanted harmful behavior is to then create a clear definition of power system attack. That would lay the basis to search for countermeasures as well.

We trained the attacker agent on a model of the power-economic system as environment, which makes the methodology only feasible for the grid operator but not the potential attacker. In future research, we want do investigate if training and successful attacks are possible without knowledge of the model. From the grid operator perspective, potential countermeasures need to investigated together with the respective attack strategy. For example, the concept of ARL mentioned in the "Related work" section lends itself here by adding a defender agent that learns defensive strategies in parallel with the attacker.

Training information

Hyperparameters TD3

The learning rate for actor and critic were both chosen as 10^{-3} with Adam as optimizer. For exploration, Gaussian noise with a standard deviation of 0.1 was added to the actions. The noise was clipped to the range [-0.5, 0.5]. The replay batch size was chosen as 100. Overall, the agents were trained for 200 tsd. steps. The actual training started after the first 10^3 samples were collected. The replay buffer had a size of 10^5.

Random seeds

All random number generators were initially seeded with noise taken from the system's `/dev/urandom` device. This ensured that even subsequent runs had sufficiently non-correlated initial seeding. For reproducibility, each run's initial seeds were saved and can be given to re-produce a specific run.

Acknowledgements
Many thanks to Martin Tröschel for internal reviewing and to Tim Hamp for valuable preliminary work.

About this supplement
This article has been published as part of *Energy Informatics Volume 4 Supplement 3, 2021: Proceedings of the 10th DACH+ Conference on Energy Informatics*. The full contents of the supplement are available online at https://energyinformatics.springeropen.com/articles/supplements/volume-4-supplement-3.

Authors' contributions
TW contributed most of the implementation, experimentation, and writing of this manuscript. EV contributed significantly to the writing of the Related Work and Reinforcement Learning sections and to the implementation as well as experimentation. AN contributed significantly to the writing of the Introduction and Related Work sections, developed the original idea, and preliminary work. All authors read and approved the final manuscript.

Funding
This work was funded by the German Federal Ministry of Education and Research through the project PYRATE (01IS19021A). Publication funding was provided by the German Federal Ministry for Economic Affairs and Energy.

Availability of data and materials
All source code and data for the experiments can be found here: https://gitlab.com/digitalized-energy-systems/scenarios/reactive_market_attack

Declarations

Competing interests
The authors declare that they have no competing interests.

Published online: 13 September 2021

References

Amjady N, Rabiee A, Shayanfar HA (2010) Pay-as-bid based reactive power market. Energy Convers Manag 51(2):376–381. https://doi.org/10.1016/j.enconman.2009.10.012

Buchholz S, Tiemann PH, Wolgast T, Scheunert A, Gerlach J, Majumdar N, Breitner M, Hofmann L, Nieße A, Weyer H (2021) A sketch of unwanted gaming strategies in flexibility provision for the energy system. In: 16th International Conference on Wirtschaftsinformatik, Pre-Conference Community Workshop Energy Informatics and Electro Mobility ICT

Chen Y, Tan Y, Zhang B (2019) Exploiting Vulnerabilities of Load Forecasting Through Adversarial Attacks. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems - e-Energy '19. ACM Press, New York, USA. pp 1–11. https://doi.org/10.1145/3307772.3328314

de Souza ACZ, Alvarado F, Glavic M (2001) The effect of loading on reactive market power. In: Sprague RH (ed). Proceedings of the 34th Annual Hawaii International Conference on System Sciences. IEEE Computer Society, Los Alamitos, Calif. https://doi.org/10.1109/HICSS.2001.926287

E-ISAC (2016) Analysis of the Cyber Attack on the Ukrainian Power Grid: Defense Use Case. Electr Inf Sharing Anal Center (E-ISAC)

Farraj A, Hammad E, Daoud AA, Kundur D (2016) A game-theoretic analysis of cyber switching attacks and mitigation in smart grid systems. IEEE Trans Smart Grid 7(4):1846–1855. https://doi.org/10.1109/TSG.2015.2440095

Fischer L, Memmen J-M, Veith EM, Tröschel M (2018) Adversarial Resilience Learning - Towards Systemic Vulnerability Analysis for Large and Complex Systems. https://arxiv.org/pdf/1811.06447

Fujimoto S, Van Hoof H, Meger D (2018) Addressing function approximation error in actor-critic methods. In: 35th International Conference on Machine Learning, ICML 2018 Vol. 4. pp 2587–2601. http://arxiv.org/abs/1802.09477

Gils HC (2014) Assessment of the theoretical demand response potential in Europe. Energy 67:1–18. https://doi.org/10.1016/j.energy.2014.02.019

Henderson P, Islam R, Bachman P, Pineau J, Precup D, Meger D (2018) Deep reinforcement learning that matters. In: Proceedings of the AAAI Conference on Artificial Intelligence Vol. 32. https://ojs.aaai.org/index.php/AAAI/article/view/11694

Hirth L, Schlecht I (2019) Market-based redispatch in zonal electricity markets: Inc-dec gaming as a consequence of inconsistent power market design (not market power). Technical report, Kiel, Hamburg. More recent version: http://hdl.handle.net/10419/194292

Hirth L, Schlecht I, Maurer C, Tersteegen B (2018) Zusammenspiel von Markt und Netz im Stromsystem: Eine Systematisierung und Bewertung von Ausgestaltungen des Strommarkts. Bundesministerium für Wirtschaft und Energie (BMWi)

Jay D, Swarup KS (2021) A comprehensive survey on reactive power ancillary service markets. Renew Sustain Energy Rev 144:110967. https://doi.org/10.1016/j.rser.2021.110967

Ju P, Lin X (2018) Adversarial attacks to distributed voltage control in power distribution networks with DERs. In: Proceedings of the Ninth International Conference on Future Energy Systems. Association for Computing Machinery, New York. pp 291–302. https://doi.org/10.1145/3208903.3208912

Kiedanski D, Orda A, Kofman D (2019) The effect of ramp constraints on coalitional storage games. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems. Association for Computing Machinery, New York, USA. pp 226–238. https://doi.org/10.1145/3307772.3328300

Konstantinidis C, Strbac G (2015) Empirics of intraday and real-time markets in europe: Great britain. Technical report, DIW – Deutsches Institut für Wirtschaftsforschung, Berlin, Germany. https://www.econstor.eu/bitstream/10419/111266/1/Report_1st_FPM_2015_UK.pdf

Liang G, Zhao J, Luo F, Weller SR, Dong ZY (2017) A Review of False Data Injection Attacks Against Modern Power Systems. IEEE Trans Smart Grid 8(4):1630–1638. https://doi.org/10.1109/TSG.2015.2495133

Minniti S, Haque N, Nguyen P, Pemen G (2018) Local Markets for Flexibility Trading: Key Stages and Enablers. Energies 11(11):3074. https://doi.org/10.3390/en11113074

Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M (2013) Playing atari with deep reinforcement learning. https://arxiv.org/pdf/1312.5602

Ni Z, Paul S, Zhong X, Wei Q (2018) A reinforcement learning approach for sequential decision-making process of attacks in smart grid. In: 2017 SSCI Proceedings. IEEE, Piscataway, NJ. pp 1–8. https://doi.org/10.1109/SSCI.2017.8285291

Paul S, Ni Z (2017) Vulnerability analysis for simultaneous attack in smart grid security. In: 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE, Piscataway, NJ. pp 1–5. https://doi.org/10.1109/ISGT.2017.8086078

Puiutta E, Veith EMSP (2020) Explainable Reinforcement Learning: A Survey. In: Holzinger A, Kieseberg P, Tjoa AM, Weippl E (eds). Machine Learning and Knowledge Extraction. Springer International Publishing, Cham. pp 77–95

Reuters (2017) Ukrainian banks, electricity firm hit by fresh cyber attack. Reuters

Samimi A, Kazemi A, Siano P (2015) Economic-environmental active and reactive power scheduling of modern distribution systems in presence of wind generations: A distribution market-based approach. Energy Convers Manag 106:495–509. https://doi.org/10.1016/j.enconman.2015.09.070

Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, Guez A, Lockhart E, Hassabis D, Graepel T, Lillicrap T, Silver D (2019) Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. ArXiv:1–21. http://arxiv.org/abs/1911.08265

Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T, Lillicrap T, Simonyan K, Hassabis D (2017) Mastering chess and shogi by self-play with a general reinforcement learning algorithm. ArXiv. http://arxiv.org/abs/1712.01815

Silver D, Schrittwieser J, Simonyan K, Nature IA, 2017 U (2016) Mastering the game of go without human knowledge. Nature 550(7676):354

Spooner T, Savani R (2020) Robust market making via adversarial reinforcement learning. In: IJCAI International Joint Conference on Artificial Intelligence. pp 4590–4596. https://doi.org/10.24963/ijcai.2020/633

Styczynski J, Beach-Westmoreland N (2016) When the lights went out: Ukraine cybersecurity threat briefing. Booz Allen Hamilton 12:20

Thurner L, Scheidler A, Schäfer F, Menke J-H, Dollichon J, Meier F, Meinecke S, Braun M (2018) pandapower - an Open Source Python Tool for Convenient Modeling, Analysis and Optimization of Electric Power Systems. IEEE Trans Power Syst 33(6):S. 6510–6521

van Hasselt H, Guez A, Silver D (2016) Deep Reinforcement Learning with Double Q-Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol 30, Issue (1). https://ojs.aaai.org/index.php/AAAI/article/view/10295

Veith EM, Balduin S, Wenninghoff N, Tröschel M, Fischer L, Nieße A, Wolgast T, Sethmann R, Fraune B, Woltjen T (2020) Analyzing Power Grid, ICT, and Market Without Domain Knowledge Using Distributed Artificial Intelligence. In: CYBER 2020, The Fifth International Conference on Cyber-Technologies and Cyber-Systems, S. pp 86–93

Veith EM, Fischer L, Tröschel M, Nieße A (2019) Analyzing cyber-physical systems from the perspective of artificial intelligence. In: Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control. Association for Computing Machinery, New York, USA. pp 85–95. https://doi.org/10.1145/3388218.3388222

Wei L, Sarwat AI, Saad W, Biswas S (2018) Stochastic games for power grid protection against coordinated cyber-physical attacks. IEEE Trans Smart Grid 9(2):684–694. https://doi.org/10.1109/TSG.2016.2561266

Wu Y, Mansimov E, Liao S, Grosse R, Ba J (2017) Scalable trust-region method for deep reinforcement learning using Kronecker-factored approximation. ArXiv. http://arxiv.org/abs/1708.05144

Xie L, Mo Y, Sinopoli B (2010) False Data Injection Attacks in Electricity Markets. In: 2010 First IEEE International Conference on Smart Grid Communications. IEEE, Manhattan, New York, USA. pp 226–231. https://doi.org/10.1109/SMARTGRID.2010.5622048

Yan J, He H, Zhong X, Tang Y (2017) Q-Learning-Based Vulnerability Analysis of Smart Grid Against Sequential Topology Attacks. IEEE Trans Inf Forensic Secur 12(1):200–210. https://doi.org/10.1109/TIFS.2016.2607701

Zhong J, Bhattacharya K (2002) Toward a competitive market for reactive power. IEEE Trans Power Syst 17(4):1206–1215. https://doi.org/10.1109/TPWRS.2002.805025

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.