Reinforcement Learning Portfolio Optimization of Electric Vehicle Virtual Power Plants

Master Thesis

Author: Tobias Richter (Student ID: 558305)
Supervisor: Univ.-Prof. Dr. Wolfgang Ketter
Co-Supervisor: Karsten Schroer

Department of Information Systems for Sustainable Society
Faculty of Management, Economics and Social Sciences
University of Cologne

March 23, 2019

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Schriften entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Die Strafbarkeit einer falschen eidesstattlichen Versicherung ist mir bekannt, namentlich die Strafandrohung gemäß § 156 StGB bis zu drei Jahren Freiheitsstrafe oder Geldstrafe bei vorsätzlicher Begehung der Tat bzw. gemäß § 161 Abs. 1 StGB bis zu einem Jahr Freiheitsstrafe oder Geldstrafe bei fahrlässiger Begehung.

Tobias Richter

Köln, den 01.05.2019

Contents

List of Figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network

DP Dynamic Programming

DSO Distribution System Operator

EPEX European Power Exchange

EV Electric Vehicle

GCRM German Control Reserve Market

GP Genetic Programming

MAW Mean Asymmetric Weighted Objective Function

MC Monte Carlo

MDP Markov Decision Process

PDF Probability Density Function

RES Renewable Energy Sources

RL Reinforcement Learning

TD Temporal-Difference

TSO Transmission System Operator

V2G Vehicle-to-Grid

VPP Virtual Power Plant

Summary of Notation

Capital letters are used for random variables, whereas lower case letters are used for the values of random variables and for scalar functions. Quantities that are required to be real-valued vectors are written in bold and in lower case (even if random variables).

$\doteq$	equality relationship that is true by definition
$\approx$	approximately equal
$\mathbb{E}[X]$	expectation of a random variable X, i.e., $\mathbb{E}[X] \doteq \sum_x p(x)x$
$\mathbb{R}$	set of real numbers
$\leftarrow$	assignment
ε	probability of taking a random action in an ε-greedy policy
α	step-size parameter
γ	discount-rate parameter
λ	decay-rate parameter for eligibility traces
s, s'	states
a	an action
r	a reward
$\mathcal{S}$	set of all nonterminal states
$\mathcal{A}$	set of all available actions
$\mathcal{R}$	set of all possible rewards, a finite subset of $\mathbb{R}$
$\subset$	subset of; e.g., $\mathcal{R} \subset \mathbb{R}$
$\in$	is an element of; e.g., $s \in \mathcal{S}$, $r \in \mathcal{R}$
t	discrete time step
$T, T(t)$	final time step of an episode, or of the episode including time step t
A_t	action at time t
S_t	state at time t, typically due, stochastically, to S_{t-1} and A_{t-1}
R_t	reward at time t, typically due, stochastically, to S_{t-1} and A_{t-1}
π	policy (decision-making rule)
$\pi(s)$	action taken in state s under *deterministic* policy π
$\pi(a\|s)$	probability of taking action a in state s under *stochastic* policy π
G_t	return following time t
$p(s', r \mid s, a)$	probability of transition to state s' with reward r, from state s and action a
$p(s' \mid s, a)$	probability of transition to state s', from state s taking action a
$v_\pi(s)$	value of state s under policy π (expected return)

$v_*(s)$	value of state s under the optimal policy
$q_\pi(s,a)$	value of taking action a in state s under policy π
$q_*(s,a)$	value of taking action a in state s under the optimal policy
V, V_t	array estimates of state-value function v_π or v_*
Q, Q_t	array estimates of action-value function q_π or q_*
d	dimensionality—the number of components of $\mathbf{w}$
$\mathbf{w}$	d-vector of weights underlying an approximate value function
$\hat{v}(s,\mathbf{w})$	approximate value of state s given weight vector $\mathbf{w}$
$\mu(s)$	on-policy distribution over states
$\overline{\text{VE}}$	mean square value error

1 Model: FleetRL

The following chapter will introduce the model of this research. In its essence, we propose a solution for EV fleet providers to utilize a VPP portfolio to profitably provide balancing services to the grid on multiple markets. A control mechanism procures energy from electricity markets, allocates available EVs to VPPs, and intelligently dispatches EVs to charge the acquired amount of energy. The model uses a RL agent that learns an optimal bidding strategy by interacting with the electricity markets and reacts to changing rental demand of the EV fleet. This chapter is structured as follows: The information assumptions are listed first, the control mechanism is explained next, and finally the RL approach is described in detail. For a table of notation refer to Table 1.

We formulate the problem as a *controlled EV charging* problem. The EV fleet operator represents the *controller*, which aims to charge the fleet at minimal costs. First, the controller predicts the amount of energy it can charge in a given *market period* h. The length of the market period Δh and the market closing time depend on the considered electricity market. Second, the controller places bids on one or multiple markets to procure the predicted amount of energy. Lastly, at electricity delivery time, the controller communicates with the EV fleet to control the charging in real-time. Online EV *control periods* t are typically shorter than market periods. In the empirical case that we consider, the market periods are 15 minutes long, while the EV control periods last 5 minutes. Nonetheless, the presented approach generalizes to other period lengths. During each control period, the controller has to take decisions which individual EVs it should dispatch to charge the procured amount of electricity. In times of unforeseen rental demand, this decision implies trading off commitments to the markets with compromising customer mobility by refusing customer rentals.

Table 1: Table of Notation

Symbol	Description	Unit
t	Control period.	-
h	Market period.	-
T	Number of control periods in a market period.	-
H	Number of market periods in day.	-
N_h	Total number of market periods.	-
Δt	Length of control period.	hours
Δh	Length of market period.	hours
P_h^{bal}	Amount of balancing power offered on the balancing market.	kW

Continued on next page

Continued from previous page

Symbol	Description	Unit
p_h^c	Capacity price offered on the balancing market.	$\frac{€}{\text{MW}}$
p_h^e	Energy price offered on the balancing market.	$\frac{€}{\text{MWh}}$
$\overline{p}_h^c$	Critical capacity price in market period h.	$\frac{€}{\text{MW}}$
$\overline{p}_h^e$	Critical energy price in market period h.	$\frac{€}{\text{MWh}}$
P_h^{intr}	Amount of power offered for the unit on the intraday market.	kW
p_h^u	Unit price offered on the intraday market.	$\frac{€}{\text{MWh}}$
$\overline{p}_h^u$	Critical unit price in market period h.	$\frac{€}{\text{MWh}}$
E_h^{bal}	Amount of energy charged from balancing market in market period h.	MWh
E_h^{intr}	Amount of energy charged from the intraday market in market period h.	MWh
p^i	Industry tariff	$\frac{€}{\text{kWh}}$
P_t^{fleet}	Amount of available fleet charging power in control period t.	kW
$\widehat{P}_t^{fleet}$	Predicted amount of available fleet charging power in control period t.	kW
$C_h^{bal}(P)$	Cost function for procuring electricity from the balancing market.	€
$C_h^{intr}(P)$	Cost function for procuring electricity from the intraday market.	€
$\rho_{t,i}$	Opportunity costs of lost rental of EV i in control period t.	€
β_h	Imbalance costs in market period h.	€
λ_h^{bal}	Balancing market risk factor.	-
λ_h^{intr}	Intraday market risk factor.	-
θ_λ	Set of risk factors for all market periods $h \in \{1, ..., N_h\}$.	-
$C^{fleet}(\theta_\lambda)$	Cost function for the fleets total costs over all market periods h.	€
C_h^{fleet}	Total accumulated fleet costs at market period h.	€
i	Electric Vehicle.	-
$\mathbf{c}_i$	Dummy variable if EV is connected to a charging station.	0/1
ω_i	Amount of electricity stored in EV.	kWh
Ω	Maximum battery capacity of EV.	kWh
δ	Charging power of EV at the charging station.	kW
$\mathcal{F}$	Set of all EVs in the fleet	-

Continued on next page

Continued from previous page

Symbol	Description	Unit
$\|\mathcal{F}\|$	Total number of EVs in the fleet.	-

1.1 Required Information Assumptions

The following information is assumed to be available:

1. The controller is able to forecast the mobility demand of the EV fleet at different time-horizons based on historical data. More specifically, it can predict the amount of plugged-in EVs and consequently the available charging power P_t^{fleet} of the fleet at control period t. The prediction's accuracy is increasing with shorter time horizons, from uncertain predictions one week ahead to very accurate predictions 30 minutes ahead. Past research presented successful mobility demand forecast algorithms in the context of free-float carsharing (Kahlen et al., 2018, 2017; Wagner et al., 2016).
2. The controller is able to forecast electricity prices of spot and balancing markets based on historical data. More specifically, it can estimate the critical prices $\overline{p}_h^c$, $\overline{p}_h^e$, and $\overline{p}_h^u$ for each market period with perfect accuracy. The critical prices form an essential piece of information for the proposed bidding strategy; bids equal or below the critical price will get accepted and result in successful electricity procurement. Electricity price forecasting is an extensively studied research area, with well-advanced prediction algorithms (Weron, 2014; Avci et al., 2018).

We are confident that taking the above assumptions is viable, assuming available forecasting information is common practice in the VPP and EV fleet charging literature, see e.g.: Brandt et al. (2017); Vandael et al. (2015); Mashhour and Moghaddas-Tafreshi (2011); Tomić and Kempton (2007); Pandžić et al. (2013).

1.2 Control Mechanism

The central control mechanism constitutes the core of this research. It can be seen as a decision support system that can be deployed at a EV fleet operator to control the charging its fleet. Figure 1 depicts the control mechanism, which is divided into three distinct phases:

The first phase, *Bidding Phase I*, takes place just before the closing time of the balancing market, once every week (e.g., Wednesdays at 3pm at the GCRM). In this phase, the controller can place bids for every market period h of the following week on the balancing market. The second phase, *Bidding Phase II*, takes places in every market period of $\Delta h = 15$ minutes. At this point, the

controller has the opportunity to place bids for the market period 30 minutes ahead. By submitting bids 30 minutes ahead of time, the controller assures that the bid will be matched until the lead time of the market (e.g, 5 minutes on the EPEX Spot Intraday Continuous). The third phase, *Dispatch Phase*, takes places in every control period of $\Delta t = 5$ minutes. In this phase the controller has to dispatch available EVs to charge the procured electricity from the markets. This phase involves allocating individual EVs to the VPP and eventually refusing customer rentals to assure that all commitments can be fulfilled.

The following chapters will highlight the important parts of the various phases and provide detailed explanation and mathematical formulations.

1.2.1 Fleet Charging Power Prediction

In a first step, the controller has to predict the available fleet charging power $\widehat{P}_h^{fleet}$ for all market periods h of the next week. The actual available fleet charging power P_t^{fleet} in a control period t is given by the number of EVs that are connected to a charging station, with enough free battery capacity to charge the next control period $t+1$. As mentioned in the Chapter 1.1, the controller is able to predict the available fleet charging power $\widehat{P}_t^{fleet}$ for all control periods t with different levels of accuracy dependent on the time horizon of t.

When the controller procures electricity from the markets, the fleet has to charge with the committed charging power during all T control periods of the market period h. To minimize the risk of not being able to charge the committed amount of energy during the whole market period, and consequently causing imbalance costs, the predicted fleet charging power in a market period is defined as the minimal predicted fleet charging power of all T control periods in a market period.

$$\widehat{P}_h^{fleet} \doteq \min_{n \in \{1,..,T\}} \widehat{P}_{t+n}^{fleet} \ , \tag{1}$$

where h is the market period of interest and t its first control period.

1.2.2 Market Decision

In a second step, the controller has to decide from which market it should procure the desired amount of energy. Therefore, it compares the costs for charging electricity from the balancing market and the intraday market. The cost function for charging electricity from the balancing market is defined as follows:

$$\begin{aligned} C_h^{bal}(P) &\doteq -(P \times 10^{-3} \times \overline{p}_h^c) + (E_h^{bal} \times \overline{p}_h^e) \\ &= -(P \times 10^{-3} \times \overline{p}_h^c) + (P \frac{\Delta h}{10^3} \times \overline{p}_h^e) \ , \end{aligned} \tag{2}$$

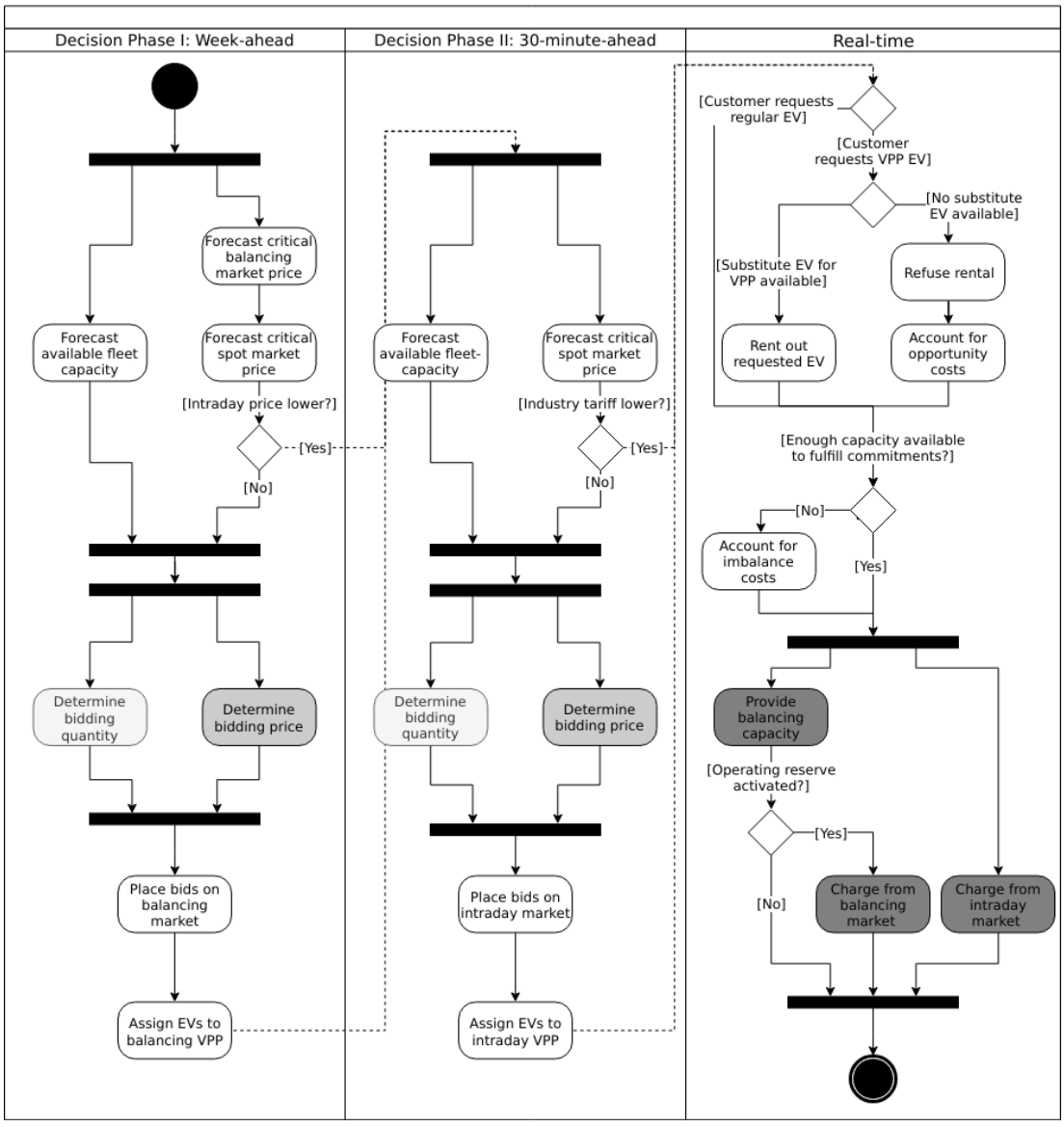

Figure 1: EV VPP Control Mechanism

where P (kW) is the amount of offered balancing power. The first term of the equation corresponds to the compensation the controller retrieves for keeping the balancing capacity available, while the second term corresponds to the costs for charging the activated balancing energy E_h^{bal} (MWh). Energy is power over time, hence E_h^{bal} can be substituted with P times the market periods length Δh, divided by the unit conversion from kW to MW. As mentioned in the Chapter 1.1, the critical prices $\overline{p}^c, \overline{p}^e, \overline{p}^u$ are assumed to be available for all market periods. Note that the critical energy price $\overline{p}^e \in \mathbb{R}$, can also take negative values, resulting in profits for the fleet, while the critical capacity price $\overline{p}^c \in \mathbb{R}_0^+$ can not take negative values and therefore always results in profits for the fleet.

The cost function for charging from the intraday market is defined similarly:

$$\begin{aligned} C_h^{intr}(P) &\doteq E_h^{intr} \times \overline{p}_h^u \\ &= P\frac{\Delta h}{10^3} \times \overline{p}_h^u \end{aligned} \tag{3}$$

Again, depending on the market situation, $\overline{p}^u \in \mathbb{R}$ can be either negative or positive, resulting in costs or profits for the fleet. Contrarily to the balancing market, on the intraday market the fleet does not get compensated for keeping the charging power available; only the charged energy affects the costs. If the costs for charging from the balancing market 7 days ahead $C^{bal}_{h+(7\times H)}(\widehat{P}^{fleet}_{h+(7\times H)})$ are higher than the costs of charging from the intraday market at the same market period $C^{intr}_{h+(7\times H)}(\widehat{P}^{fleet}_{h+(7\times H)})$, the controller does not place bids on the balancing market.

1.2.3 Determining the Bidding Quantity

In a third step, the controller has to take a decision on the amount of energy it should procure from the markets. Determining the bidding quantity is a central piece of the controlled charging problem. The bidding quantity determines the profits that can be made, by charging at a cheaper market price than the flat industry tariff. In order to maximize its profits, the controller aims to procure as much electricity as possible from the markets. In order to optimally place bids on the electricity markets, it needs to balance the risk of (a) procuring more energy that it can maximally charge and (b) not procuring enough energy from the market to sufficiently charge the fleet.

In the first case (a), the fleet is facing costs of compromising customer mobility, or worse, high imbalance penalties from the markets. Renting out EVs is considerably more profitable than using EVs as a VPP to participate on the electricity markets. Refusing customer rentals, in order to fulfill market commitments, induces opportunity costs of lost rentals ρ on the fleet. Imbalance costs β

occur, when the fleet can not charge the committed amount energy at all, even with refusing rentals. In the second case (b), the fleet also faces opportunity costs of lost rentals when individual EVs do now have enough SoC for planned trips of arriving customers.

The controller faces additional risks by bidding one week ahead on the balancing market, in contrast to only 30 minutes ahead on the intraday market, as the predictions of available charging power are more uncertain with the larger time horizon. To account for all the mentioned risks, we introduce a *risk factor* $\lambda \in \mathbb{R}_{0\leq\lambda\leq1}$, where $\lambda = 0$ indicates no risk, and $\lambda = 1$ indicates a high risk. The controller determines the bidding quantity P_h^{bal} by discounting the predicted available fleet charging power $\widehat{P}_h^{fleet}$ with the possible risk λ_h of imbalance or opportunity costs:

$$P_h^{bal} \doteq \begin{cases} 0, & \text{if } C_h^{bal}(\widehat{P}_h^{fleet}) \geq E_h^{bal} 10^3 \times p^i \\ 0, & \text{if } C_h^{bal}(\widehat{P}_h^{fleet}) \geq C_h^{intr}(\widehat{P}_h^{fleet}) \\ \widehat{P}_h^{fleet} \times (1-\lambda_h^{bal}), & \text{otherwise} \end{cases} \tag{4}$$

where h is the market period of interest one week ahead. If the controller can buy electricity at the intraday market at a lower price, it does not place a bid at the balancing market. If the controller can charge cheaper at the regular industry tariff p^i, it does not place a bid either. In all other cases, the controller submits P_h^{bal} to the market.

The bidding quantity for the intraday market P_h^{intr} depends on the previously committed charging power P_h^{bal} and the newly predicted charging power $\widehat{P}_h^{fleet}$:

$$P_h^{intr} \doteq \begin{cases} 0, & \text{if } C_h^{intr}(\widehat{P}_h^{fleet} - P_h^{bal}) \geq E_h^{intr} 10^3 \times p^i \\ (\widehat{P}_h^{fleet} - P_h^{bal}) \times (1-\lambda_h^{intr}), & \text{otherwise} \end{cases} \tag{5}$$

where h is the market period of interest 30 minutes ahead. Note that any amount of electricity that the controller procured from the balancing market, does not need to be bought from intraday market for the same market period. Since the predicted charging power $\widehat{P}_h^{fleet}$ is expected to be more accurate 30 minutes ahead than one week ahead, the controller is able to correct bidding errors it made in the first decision phase, and optimally charge the whole EV fleet.

1.2.4 Dispatching Electronic Vehicle Charging

In the last step, at electricity delivery time, the EVs have to be assigned to the VPP and be *dispatched* to charge. Therefore the controller first needs to detect how many EVs are eligible to be used as VPP per control period t. EVs are

eligible if they (a) are connected to a charging station ($\mathbf{c}_i$), and (b) have enough free battery storage available ($\Omega - \omega_i$) to charge the next control period. Hence, the VPP is defined as:

$$VPP \doteq \{i \mid i \in \mathcal{F} \vee \mathbf{c}_i = 1 \vee \Omega - \omega_i \geq \gamma \Delta t\} \ , \tag{6}$$

where $\gamma \Delta t$ (kWh) denotes the amount of energy that can be charged with the charging speed of γ (kW) in control period t.

Remember that the fleet has to provide the committed charging power $P_h^{bal} + P_h^{intr}$ across all control periods t of the market period h, independent of which individual EVs are actually charging the electricity. This fact allows the controller to dynamically dispatch EVs every control period and react to unforeseen rental demand. If a customer want to rent out an EV that is assigned to the VPP, the controller only has to refuse the rental, if no other EV is available to charge instead. When no replacement EV is available, the controller has to account for lost rental profits $\rho_{t,i}$. If the VPPs total amount of available charging power $|VPP| \times \gamma$ is not sufficient to provide the total market commitments $P_h^{bal} + P_h^{intr}$, the fleet gets charged imbalance costs β_h. Otherwise all the committed energy can be charged by the VPP.

1.2.5 Evaluating the Bidding Risk

The controllers central goal is to choose the risk factors λ_h^{bal}, λ_h^{intr} for every market period h, that minimize the cost of charging, while avoiding the risks of lost rental profits $\rho_{t,i}$ or imbalance costs β_h. The total fleet costs are defined as follows:

$$C^{fleet}(\theta_\lambda) \doteq \sum_h^{N_h} \left[C_h^{bal}(P_h^{bal}) + C_h^{intr}(P_h^{intr}) + \beta_h + \sum_t^T \sum_i^{|\mathcal{F}|} \rho_{t,i} \right] \ , \tag{7}$$

where $\theta_\lambda \in \mathbb{R}^{2 \times N_h}_{0 \leq \lambda \leq 1}$ is the matrix of the risk factors λ_h^{bal}, λ_h^{intr} for all considered market periods N_h. $\mathcal{F}$ denotes the set of all EVs i in the fleet and $|\mathcal{F}|$ the fleet size. The costs for charging $C_h^{bal}(P_h^{bal})$, $C_h^{intr}(P_h^{intr})$ are clearly dependent on the chosen risk factors λ_h^{bal}, λ_h^{intr} (see Eq. 4, Eq. 5). In summary, the problem can be formulated as minimizing the total costs of the fleet, by choosing the optimal set of risk factors θ_λ:

$$\begin{aligned} \underset{\theta_\lambda}{\text{minimize}} \quad & C^{fleet}(\theta_\lambda) \\ \text{subject to} \quad & 0 \leq \lambda_h^{bal} \leq 1, \ \forall \lambda_h^{bal} \in \theta_\lambda \\ & 0 \leq \lambda_h^{intr} \leq 1, \ \forall \lambda_h^{intr} \in \theta_\lambda \end{aligned} \tag{8}$$

Solving this optimization problem with common methods like stochastic programming is a difficult task, assuming that complete information of available charging power and future electricity market prices is not always available. Since one goal of this research is to develop a model that can be applied to previously unknown settings and learn from uncertain environments, as mobility and electricity markets, we chose to solve the problem with a RL learning approach that is explained in detail in Chapter 1.3.

1.2.6 Example

At 3pm on the 9$^{\text{th}}$ of August 2017, the controller enters the first bidding phase for procuring electricity one week ahead, the market period h = *16.08.2017 15:00-15:15*. It predicts that at that point in time 250 EVs are connected to a charging station, resulting in 900kW available fleet charging power ($\widehat{P}_h^{fleet} = 900\text{kW}$), given the charging power of 3.6kW per EV. Assuming the available critical prices $\overline{p}_h^c = 5\frac{€}{\text{MW}}$, $\overline{p}_h^e = -10\frac{€}{\text{MWh}}$, and $\overline{p}_h^u = 10\frac{€}{\text{MWh}}$ for that market period, the controller now evaluates the cheapest charging option. The flat industry electricity tariff is assumed to be $p_i = 0.15\frac{€}{\text{kWh}}$. The costs for charging with the maximal amount of power $\widehat{P}_h^{fleet}$ from the balancing market ($C_h^{bal}(900\text{kW}) = -6.25€$) are less than charging from the intraday market ($C_h^{intr}(900\text{kW}) = 2.25€$) or charging at the industry tariff ($900\text{kW} \times 0.25\text{h} \times 0.15\frac{€}{\text{kWh}} = 33.75€$). In this example, by choosing the cheapest option, the balancing market, the fleet operator will even get compensated for charging its fleet.

In the next step, the controller has to submit bids to the balancing market. The RL agent determined that the risk of bidding on the balancing market is $\lambda_h^{bal} = 0.3$. Consequently, the controller sets the bidding quantity to $P_h^{bal} = \widehat{P}_h^{fleet} \times (1 - \lambda_h^{bal}) = 900\text{kW} \times 0.7 = 630\text{kW}$ and submits a bid to the market. Since we are assuming that bids at the critical price, will always get accepted, the controller procures 630kW from the balancing market and updates its account with $C_h^{bal}(630\text{kW}) = -4.725€$.

One week later, at 2:30pm on the 16$^{\text{th}}$ of August 2017, the controller enters the second bidding phase. With a time horizon of 30 minutes, it predicts less available fleet charging power of $\widehat{P}_h^{fleet} = 810\text{kW}$ for the same market period *16.08.2017-15:00*. By trading at the intraday market, the controller can now charge the remaining available EVs with a low risk of procuring more energy than it can maximally charge. At this point in time, the RL agent determines a remaining risk of $\lambda_h^{intr} = 0.05$, and sets the bidding quantity to $P_h^{intr} = (810\text{kW} - 630\text{kW}) \times (1 - 0.05) = 171\text{kW}$. Hence, the controller procures 171kW from the intraday market and updates its account with $C_h^{intr}(171\text{kW}) = 0.4275€$.

At electricity delivery time, the 16$^{\text{th}}$ of August 2017 at 3:00pm, the controller

detects 255 available EVs; EVs which are connected to a charging station and have enough battery capacity left to be charged in the next control period. It assigns 223 EVs to provide the committed 801kW charging power for the market period time Δh of 15 minutes. During that time, three customers want to rent out EVs that are allocated to the VPP. The first two rentals are accepted, because two other EVs are available to charge instead. The third rental has be to refused, since no EV is remaining as substitution. The controller has to account for the opportunity costs of the lost rental $\rho_{t,i}$.

1.3 Reinforcement Learning Approach

In the following chapter the developed RL approach is outlined. First, the we define the previously explained charging problem as a MDP, second, the used learning algorithm is explained, and third, we give implementation and training details.

Remember that the goal of the controlled charging problem is to choose a set of risk factors θ_λ that minimize the fleets total costs across all market periods. The controllers can influencing the charging costs, by setting the risk factors λ^{intr}, λ^{bal}, which determine the bidding quantities P^{bal}, P^{intr}. The controller has the possibility to submit bids to the markets every market period h. Hence, the RL agent has to take an action (i.e., determining the risk factors) every timestep of $t = h$.

The agent does so, by learning a policy $\pi(a|s)$ that maps every state $s \in \mathcal{S}$ to an action $a \in \mathcal{A}$. After the training period, the agent takes action a whenever it observes a state $S_t \in \mathcal{S}$.

1.3.1 Markov Decision Process Definition

- $a \in A$
- Both bidding risk dependent on how many EVs available at that point in time
- Both bidding risk dependent on daytime in hours at that point in time
- Intraday Bidding risk dependent on how many EVs are available now

$$|\widehat{VPP}|_t \doteq \left\lceil \frac{\widehat{P}_t^{fleet}}{\gamma} \right\rceil \tag{9}$$

The state space is modeled with the dimensions 1) Predicted charging power week-ahead, 2) time of the day of market period h. The risk of bidding on the markets is expected to be

$$\mathcal{S} \doteq \left\{ h(t), |VPP|_t, |\widehat{VPP}|_{t+2}, |\widehat{VPP}|_{t+(7\times H)} \right\} \tag{10}$$

where:

- $h(t)$ is the current daytime in hours, with discrete values in the range $\big[0,\ 23\big] \in \mathbb{N}$.
- $|VPP|_t$ is the current VPP size, with discrete values in the range $\big[0,\ |\mathcal{F}|\big] \in \mathbb{N}$.
- $|\widehat{VPP}|_{t+2}$ is the predicted VPP size 30 minutes ahead, with discrete values in the range $\big[0,\ |\mathcal{F}|\big] \in \mathbb{N}$.
- $|\widehat{VPP}|_{t+(7\times H)}$ is the predicted VPP size 7 days ahead, with discrete values in the range $\big[0,\ |\mathcal{F}|\big] \in \mathbb{N}$.

  Results in $|\mathcal{F}|^3 \times 24 = 3 \times 10^9$ states.
- See (Reddy & Veloso, 2011) for State/Action notation

  Controller takes an action every

$$\mathcal{A} \doteq \{\lambda_t^{bal}, \lambda_t^{intr}\} \tag{11}$$

where:

- λ_t^{bal} is the risk factor for bidding on the balancing market, with discrete values in the range $\big[0, 1\big]$ in 0.1 increments.
- λ_t^{intr} is the risk factor for bidding on the intraday market, with discrete values in the range $\big[0, 1\big]$ in 0.1 increments.

  Results in $10^2 = 100$ actions.

  We define the reward function as the costs that occurred in the last timestep (see Eq. 7).

- If the RL agents chose risk factors that....
- Prof

$$R = C_t^{fleet} - C_{t-1}^{fleet}\ , \tag{12}$$

where C_t^{fleet} are the total accumulated fleet costs until market period t.

1.3.2 Learning Algorithm

- (Dauer et al., 2013),
- (Di Giorgio et al., 2013)
- (Vaya et al., 2014)
- (Vandael et al., 2015)
- Cost function?
- Deep double Q learning (van Hasselt et al., 2016)
  - Image: (Wang et al., 2015)
  - double q learning combined with deep q learning:
  - off-policy approach

1.3.3 Implementation and Training

- Software:
  - Python 3.6
  - Lib: Keras-RL - Tensorflow abstraction / Pytorch?
  - Implementation Link
- Hardware: Google Colab
  - GPU: Nvidia Tesla K80 , 2880 x 2 CUDA cores, 12GB GDDR5 VRAM
  - CPU: Intel(R) Xeon(R) CPU @ 2.30GHz (1 core, 2 threads), 45MB Cache
  - RAM: ~12.6 GB Available
- Training:
  - Time
  - Hyperparameters? - In results or here?

2 Simulation Platform: FleetSim

- Simulation Platform to allow evaluating the performance of intelligent agents in smart charging/balancing the grid of EV fleets. Allows to test out bidding strategies and control mechanisms in a realistic EV fleet setting.
- Real life comparison graph: 10%.

2.1 Event-based Simulation

- Not only t+1
- Event e.g. denying rental, charge/little charge/no-charge has effects for the whole simulation
- Simpy
- Python

2.2 Architecture / Components

2.3 Modular Expandability

- Plug-in different Market designs
- Use different real-world data
- Change Fleet parameters
- Develop new strategy
-

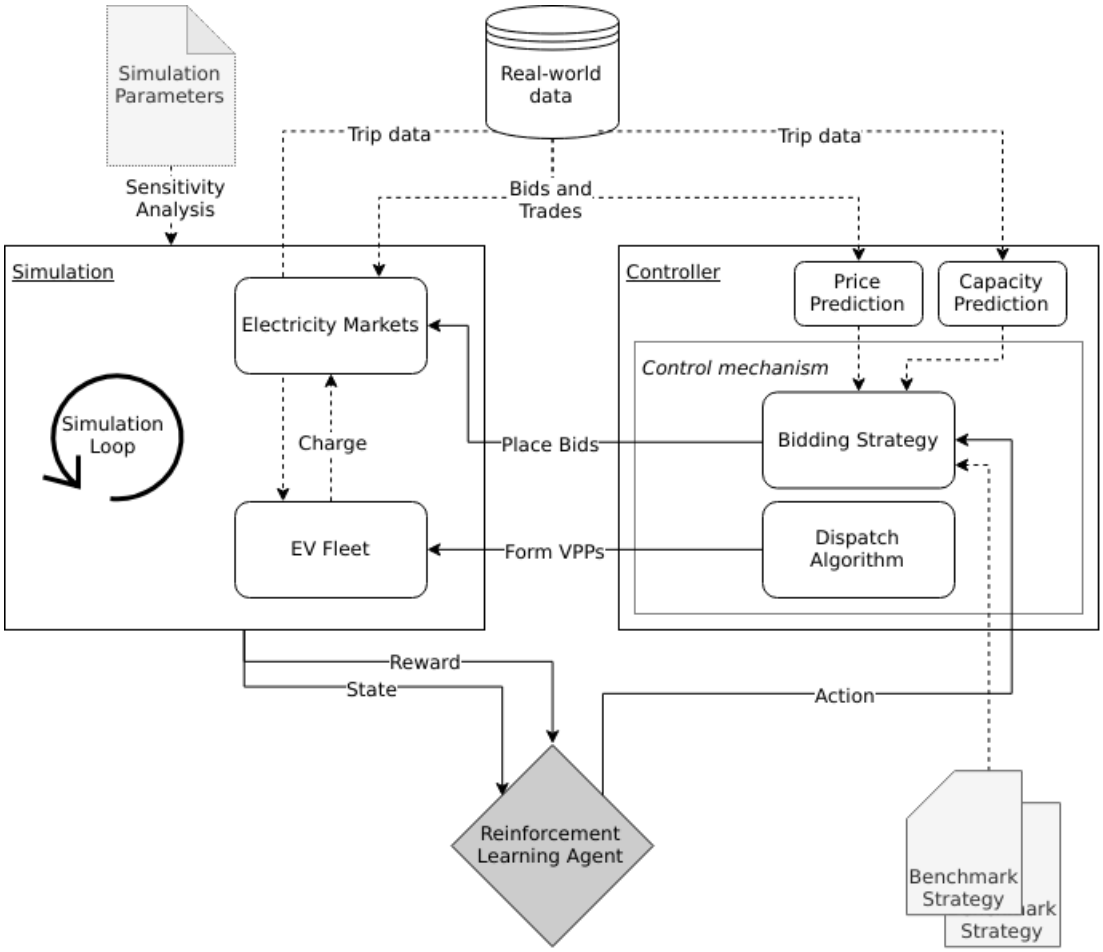

Figure 2: Architecture of FleetSim

3 Results

3.1 Simulation Settings

- Results heavily dependent on industry charging price, since on average the balancing prices are 50% cheaper, and intraday 30% cheaper.
- BMWi. (n.d.). Prices of electricity for the industry in Germany from 2008 to 2017 (in euro cents per kilowatt hour). In Statista - The Statistics Portal. Retrieved March 18, 2019, from `https://www.statista.com/statistics/595803/electricity-industry-price-germany/`.

3.2 FleetRL

- long-delayed rewards make RL hard (!?)

3.3 Sensitivity Analysis

- Prediction Accuracy
- Charging infrastructure

4 Conclusion

4.1 Contribution

- Compare to most simular studies:

(Kahlen et al., 2018; Vandael et al., 2015) etc..

- Business model for EV fleet owners with better results than previous studies
- Environmental impact by providing balancing power
- Decision Support System for controlled EV charging from multiple markets
- RL Algorithm that is designed to work in previously unknown environments and thus suited to deploy in real life settings of all kinds of EV fleets in all kinds of cities. E.g. scooters also?
- Event-based Simulation Platform to evaluate bidding strategies and RL agents, facilitate research

4.2 Limitations

- Model:
  - Bidding Mechanism: one week ahead, always accepted
  - Policy & Regulation: EVs not allowed to provide balancing power, minimum bidding quantities 1MW.

  - Markets: Fleet is a price-taker, what about larger fleets? Simulate market influence
- RL: See (Vázquez-Canteli & Nagy, 2019) conclusion for limitations.

4.3 Future Research

- Model: Current market design, i.e. daily w/ 4h slots. German "Mischpreisverfahren"
- RL: Long-delayed rewards, different reward structure, memory based

References

Avci, E., Ketter, W., & van Heck, E. (2018). Managing Electricity Price Modeling Risk Via Ensemble Forecasting: the Case of Turkey. *Energy Policy*, 390-403. Retrieved from https://doi.org/10.1016/j.enpol.2018.08.053 doi: 10.1016/j.enpol.2018.08.053

Brandt, T., Wagner, S., & Neumann, D. (2017). Evaluating a Business Model for Vehicle-Grid Integration: Evidence From Germany. *Transportation Research Part D: Transport and Environment*, 488-504. Retrieved from https://doi.org/10.1016/j.trd.2016.11.017 doi: 10.1016/j.trd.2016.11.017

Dauer, D., Flath, C. M., Strohle, P., & Weinhardt, C. (2013). Market-Based EV Charging Coordination. In *Ieee/wic/acm international joint conferences on web intelligence (wi) and intelligent agent technologies (iat).* Retrieved from https://doi.org/10.1109/wi-iat.2013.97 doi: 10.1109/wi-iat.2013.97

Di Giorgio, A., Liberati, F., & Pietrabissa, A. (2013). On-board stochastic control of Electric Vehicle recharging. In *52nd ieee conference on decision and control.* Retrieved from https://doi.org/10.1109/cdc.2013.6760789 doi: 10.1109/cdc.2013.6760789

Kahlen, M., Ketter, W., & Gupta, A. (2017). Fleetpower: Creating Virtual Power Plants in Sustainable Smart Electricity Markets. *SSRN Electronic Journal*. Retrieved from https://doi.org/10.2139/ssrn.3062433 doi: 10.2139/ssrn.3062433

Kahlen, M., Ketter, W., & van Dalen, J. (2018). Electric Vehicle Virtual Power Plant Dilemma: Grid Balancing Versus Customer Mobility. *Production and Operations Management*. Retrieved from https://doi.org/10.1111/poms.12876 doi: 10.1111/poms.12876

Mashhour, E., & Moghaddas-Tafreshi, S. M. (2011). Bidding Strategy of Virtual Power Plant for Participating in Energy and Spinning Reserve Markets-Part I: Problem Formulation. *IEEE Transactions on Power Systems*, 949-956. Retrieved from https://doi.org/10.1109/tpwrs.2010.2070884 doi: 10.1109/tpwrs.2010.2070884

Pandžić, H., Morales, J. M., Conejo, A. J., & Kuzle, I. (2013). Offering Model for a Virtual Power Plant Based on Stochastic Programming. *Applied Energy*. Retrieved from https://doi.org/10.1016/j.apenergy.2012.12.077 doi: 10.1016/j.apenergy.2012.12.077

Reddy, P. P., & Veloso, M. M. (2011). Strategy learning for autonomous agents in smart grid markets. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence.*

Tomić, J., & Kempton, W. (2007). Using Fleets of Electric-Drive Vehicles for Grid Support. *Journal of Power Sources*, 459-468. Retrieved from `https://doi.org/10.1016/j.jpowsour.2007.03.010` doi: 10.1016/j.jpowsour.2007.03.010

Vandael, S., Claessens, B., Ernst, D., Holvoet, T., & Deconinck, G. (2015). Reinforcement Learning of Heuristic EV Fleet Charging in a Day-Ahead Electricity Market. *IEEE Transactions on Smart Grid*, 1795-1805. Retrieved from `https://doi.org/10.1109/tsg.2015.2393059` doi: 10.1109/tsg.2015.2393059

van Hasselt, H., Guez, A., & Silver, D. (2016). Deep Reinforcement Learning with Double Q-Learning. In *Aaai.*

Vaya, M. G., Rosello, L. B., & Andersson, G. (2014). Optimal bidding of plug-in electric vehicles in a market-based control setup. In *Power systems computation conference.* Retrieved from `https://doi.org/10.1109/pscc.2014.7038108` doi: 10.1109/pscc.2014.7038108

Vázquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement Learning for Demand Response: a Review of Algorithms and Modeling Techniques. *Applied Energy*, 1072-1089. Retrieved from `https://doi.org/10.1016/j.apenergy.2018.11.002` doi: 10.1016/j.apenergy.2018.11.002

Wagner, S., Brandt, T., & Neumann, D. (2016). In Free Float: Developing Business Analytics Support for Carsharing Providers. *Omega*, 4-14. Retrieved from `https://doi.org/10.1016/j.omega.2015.02.011` doi: 10.1016/j.omega.2015.02.011

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & De Freitas, N. (2015). Dueling Network Architectures for Deep Reinforcement Learning. *arXiv preprint arXiv:1511.06581*.

Weron, R. (2014). Electricity Price Forecasting: a Review of the State-Of-The-Art With a Look Into the Future. *International Journal of Forecasting*, 1030-1081. Retrieved from `https://doi.org/10.1016/j.ijforecast.2014.08.008` doi: 10.1016/j.ijforecast.2014.08.008