Flexible Charging Optimization for Electric Vehicles using MDPs-based Online Algorithms

Nikita V. Tomin, [*] **Jonas Maasmann** [**]
Alexandr B. Domyshev [*]

[*] *Melentiev Energy Systems Institute SB RAS, Irkutsk, 664033 Russia (e-mail: tomin.nv@gmail.com).*
[**] *Institute of Energy Systems, Energy Efficiency and Energy Economics, TU Dortmund, Dortmund, Germany, (e-mail: jonas.maasmann@tu-dortmund.de)*

Abstract: In the paper, we formulate the problem of charging electric vehicles with a time-dependent energy source as a Markov Decision Process (MDP), with states defining the presence of cars, their individual levels of charge as well as the level of available renewable energy and storage devices. We exploit MDPs-based online algorithms such as Monte-Carlo Tree Search (MCTS) to overcome the scalability issues associated with charging of a large number of EVs, which corresponds to real distributed networks with flexible options. Using MCTS, we were able to generate optimal policies that balanced the energy toll on the electric grid with the final charge levels of each vehicle. We compare the performance of offline MDP solvers (Discrete Value Iteration algorithm) and online MDP solvers (MCTS) as well as reinforcement learning-based solvers (Q-learning) to find the optimal policy for EV's flexible charging optimization.

Keywords: electric vehicles, charging, flexibility, Markov decision process, stochastic optimization

1. INTRODUCTION

The large-scale integration of electric vehicles (EVs) into the power grid brings both challenges and opportunities to the system performance. On one hand, the load demand from EV charging imposes a large impact on the stability and efficiency of the power grid. On the other hand, EVs could potentially act as mobile energy storage systems to improve power grid performance, such as load flattening, fast frequency control, and facilitating renewable energy integration. Evidently, uncontrolled EV charging could lead to inefficient power network operation or even security.

Since deep market penetration of EVs will impose substantial current demands on an already fragile electricity grid, an optimal policy is sought to schedule smooth charging of EVs overnight and minimize the need for non-renewable electricity sources. In the paper, we propose to do by structuring the problem as a Markov Decision Processes (MDP) and performing online MDP solvers

2. PROBLEM STATEMENT

Especially in the distribution network, an uncontrolled integration of EVs causes an increased need for grid expansion. Table 1 shows the estimated grid expansion costs of different studies for the German grid. They calculate costs between 11bn euro and 253bn euro for uncontrolled and uncoordinated charging of EV's battery for different scenarios.

Table 1. Estimated Grid Extension Cost

Study	Year	Grid Level	EV penetration	Grid extension costs
Pregger and et al (2012)	2012	HV, MV, LV	5,1 Mio	$\prec$ 3%
Friedl and et al. (2018)	2018	LV	50%, 100%	11, 26 bn euro
Brundlinger and et al. (2017)	2017	HV, MV, LV	100%	146-253 bn euro

Uncontrolled charging processes, especially on private charging infrastructure, cause load profiles whose maximum values are at the same time as the maximum of the standard load profile. Figure 1 shows the load profiles of the two charging cases "charging after arriving after the last trip" and "charging after arriving after work". Both profiles are for charging at home and calculated for a maximum charging power of 3.7 kW and 11 kW J.Maasmann (2019).

In combination with the standard load profile overload situations or voltage violations are the consequence. A temporal shift of the two load peaks can lead to a reduction of the cumulated total load and thus prevent overload

[⋆] This work was supported by the Russian Scientific Foundation (No. 19-49-04108) and the German Science Foundation/DFG (No. RE 2930/24) under the project "Development of Innovative Technologies and Tools for Flexibility Assessment and Enhancement of Future Power Systems".

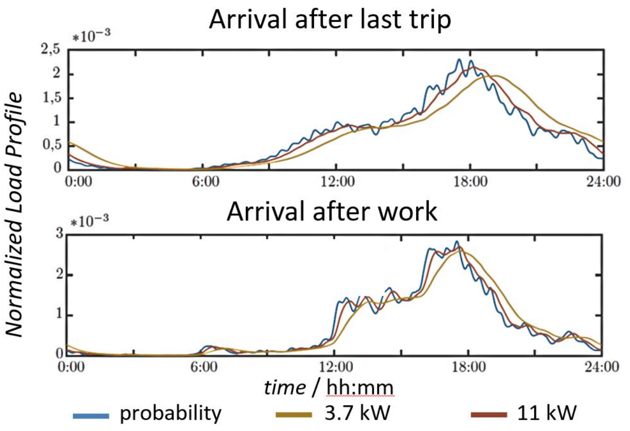

Fig. 1. Load profiles with EVs of the two charging cases

hbt

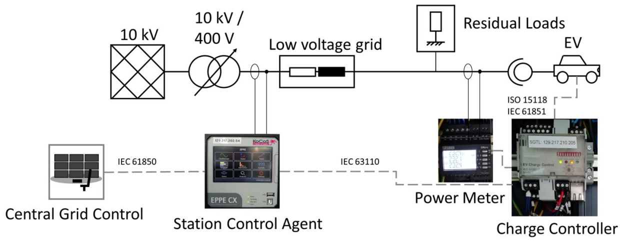

Fig. 2. The possible configuration to handle grid controlled charging

situations. This shift is made possible by making loading processes flexible. A communication infrastructure is used to determine which requirements the user of the EVs has and which grid constraints exist. An algorithm then determines a charging model, which optimizes the charging processes for the gird. In addition to the goal of preventing critical grid conditions, such charging models can also have additional optimization goals, such as the increased use of renewable energies.

The charging models presented in the technical literature, however, overlook the stochastic nature of driving patterns and grid constraints. Here we introduce an efficient stochastic model to optimally charge an EV while accounting for the uncertainty inherent to its use and considering distribution grid constraints Tang et al. (2016).

To handle grid controlled charging proceeds Figure 2 shows a possible configuration. The charging models and grid optimization runs on the central grid control. This aggregate information due to station control agents from different charge controller in the connected charging stations. After optimization the central grid control system, communicate the optimal power for the charge procedure via the station control agent to the charging controller. The charging controller adjust the maximum or optimal charge power to the EV.

A classification of handle charging problem can be separated in three general research fields. To find out more about the general behavior of charging EV with the electrical grid empiric approaches like Metz and Doetsch (2012) are necessary who analyses the behavior of the mobility. Out of these studies, the impact of to the electrical grid and the need of flexibility can be evaluated. Research on technology with applied approaches in lab and field tests shows the possibility for using controlled charging to generate flexibility during the charging process for different control strategies (e.g. Maasmann et al. (2014)).

The design of the control strategies differs with the objective of the control. A promising approach is the formulation of an optimization problem to create a schedule for the charging processes Mukherjee and Gupta (2015). Especially for optimization of minimizing financial costs Sortomme and El-Sharkawi (2012) or to reduce CO2 emission Jin et al. (2012) under constrains from electrical grid an optimization of charging schedule is helpful. Important is a scalable and efficient algorithm for solving there optimizations problems. This approach focus on MDP solvers for handle charging optimization of EV.

3. MDP DEFINITION FOR CHARGING AN EV

3.1 Markov Decision Processes

Charging requires some form of feedback from the EV so that it only is pulling energy from the grid during off-peak times, and only to the extent that the grid can sustainably handle Mwasilu et al. (2014); Hadley and Tsvetkova (2009). Therefore, since both of these ideas rely on a knowledgeable car charging energy distribution cycle, a problem wrapped in uncertainties, it becomes the basis for this research.

In order to solve this decision-making problem, it is useful to model the scenario as a MDP, which allows for principled decision making under conditions of uncertain sensing. An MDP is a mathematical framework for sequential decision making under uncertainty, and where all of the uncertainty arises from outcomes that are partially random and partially under the control of a decision maker. Mathematically, an MDP is a tuple (S, A, T, R), where S is the state space, A is the action space, T is a transition function defining the probability of transitioning to each state given the state and action at the previous time, and R is a reward function mapping every possible transition (s, a, s') to a real reward value Kochenderfer (2015).

3.2 Online and offline solvers

Sequential decision making under uncertainty involves both online and offline calculations. Methods that use MDP framework to solve planning in an imperfectly known and dynamic environment can be classified into two approaches - offline and online. The first approach embeds all possible environments and their dynamics as part of the MDP model. It uses an offline MDP solver to find a good policy (strategy), prior to execution. When the environment and its dynamics are largely unknown, this approach constructs MDP models too huge to be solved by even the best offline solver today.

The second approach (online) models only the known part of the environment and its dynamics (both stochastic and deterministic), and allows the model to change during execution when more information about the environment becomes available. The key to the success of this approach is an efficient online MDP solver that can compute a good policy during runtime, following changes in the MDP model.

For the task of optimization and planning of charging EVs, the scalability problem becomes relevant with an increase in the number of machines and other flexibility options in the distribution grid. One of the options would be to use a MDP-based online method. This would only consider the states that are reachable from the current state and would, therefore, limit the computational power and storage required for computation by again trading off the certainty of optimality. One of the popular online MDP approaches now is Monte Carlo Tree Search (MCTS). This solver became the foundation of an advanced AI system - AlphaGO Zero from the Google DeepMind Silver et al. (2017). In Li and Du (2018); Tomin et al. (2019) several potential MCTS applications in power systems were proposed, including coordinated management of plug-in EVs.

MCTS is a policy-optimization algorithm for finite-horizon, finite-size MDP, based on random episode sampling structured by a decision tree. MCTS proceeds in four phases of selection, expansion, rollout, and back-propagation. The standard MCTS algorithm proceeds by repeatedly adding one node at a time to the current tree. Given that leaf nodes are likely to be far from terminal states, it uses random actions, to estimate state-action values. After the rollout phase, the total collected rewards during the episode is back-propagated through the tree branch, updating their empirical state-action values, and visit counts. Upper Confidence Bounds (UCB) is an optimization algorithm that is used for choosing which child node to expand (i.e., choosing an action) Kartal et al. (2019). Each parent node chooses its childs with the largest $USB(s_t, a_t)$ value according to the following formula:

$$USB(s_t, a_t) = Q(s_t, a_t) + C\sqrt{(ln(N_p)/(1 + N_i)} \qquad (1)$$

where N_i is the visit count for ith child; N_p is the number of visit counts for the parent node The parameter $c \geq 0$ controls the tradeoff between choosing lucrative nodes (low c) and exploring nodes with low visit counts (high c); $Q(s_t, a_t)$ is the state-action value function associated to an optimal policy π^* is used to characterize the quality of taking action a_t at state s_t and then acting optimally and is defined as:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma min_{a_{t+1}}(Q_{t+1}(s_{t+1}, a_{t+1})) \qquad (2)$$

It's important to note that one of the popular reinforcement learning (RL) approaches is Q-learning algorithm, which implements an iterative approximation of $Q(s_t, a_t)$ through training on temporal differences, when the mean square error between the predictor and the goal is minimized at each step (see Eq. 2). RL solves the problem of sequential optimal decision making Sutton and Barto (2018). The mathematical model of this problem is MDP.

3.3 Methodology

The problem of charging an EV can be posed as a conflict between two opposing objectives. The end-user desires to have the EV charged and ready for use at his/her discretion, while also minimizing the costs of running the EV. Demand for electricity varies over the day and so does the electricity generated from renewable sources. This introduces a varying energy price which can make t beneficial for the end-user to postpone charging his/her EV. This means the user is faced with the problem of postponing charging to minimize costs or to charge right away so as to maximize the availability of the EV.

An optimal policy is sought to schedule smooth charging of EVs overnight and minimize the need for non-renewable electricity sources. This is done by structuring the problem as a MDP and performing various solvers (online and offline). We scale the MDP to the level of a distribution network with homes as typical consumers. In this case, we model n homes on the electric grid, each with an EV charging port capable of charging one car. It is assumed that each car follows a unique driving route during the day and arrives back to the home for charging at night at a variable time and with variable amount of current charge c.

In this model, the arrival times of the EVs are determined on the basis of real measured user behavior. This provides a more realistic representation of the results than is possible by using standard load profiles. The departure to work varies between seven and nine o'clock and the return home between 16 and 22 o'clock (Fig. 3 and Fig. 4). The loading outside the house is not considered for the creation of this load profile.

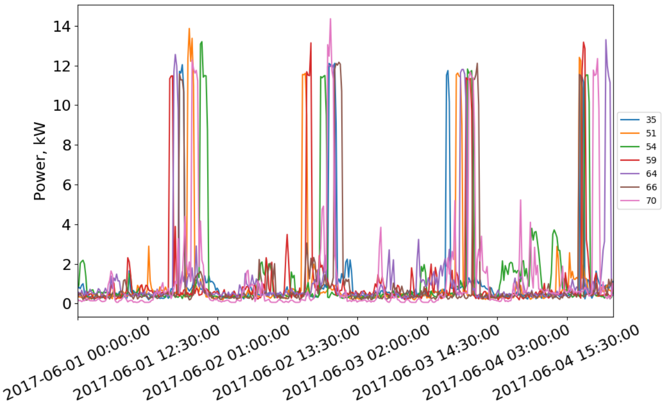

Fig. 3. The example of house load profiles with 'grid friendly' charging of EVs for several consumers of HTW Berlin

We simplify the simulation of the driving patterns of EVs and their corresponding state of charge through transition function for the MDP environment (Fig. 5). In our study, the transition model mandates that time increments by one, or $t_t = t_{t-1} + 1$. Additionally, if a EV is present, it will remain until the end of the simulation. If a EV is not present, the probability of it arriving at the next time step is

$$P_{n,t} = \frac{1}{1 + exp(-20\frac{t-1}{T})} \qquad (3)$$

When a car first arrives, it's initial charge may take any value between 0 and $C - 1$ with equal probability. If a car was present and the action to charge that car is taken, the level is incremented by one, up to a maximum charge potential C.

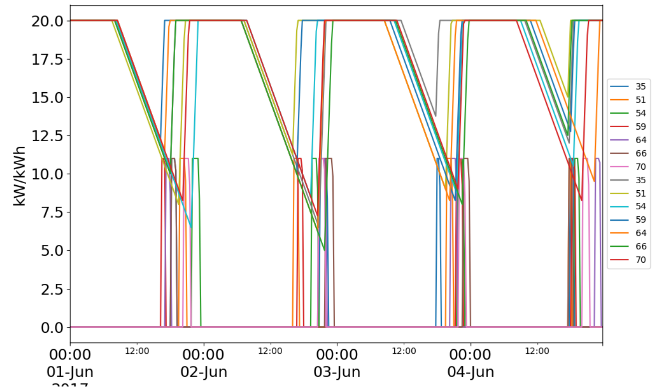

Fig. 4. Example of the charge pattern of the EV's battery. Above is the state of charge of the vehicle in kWh, below the power reference of the charging station in kW shown.

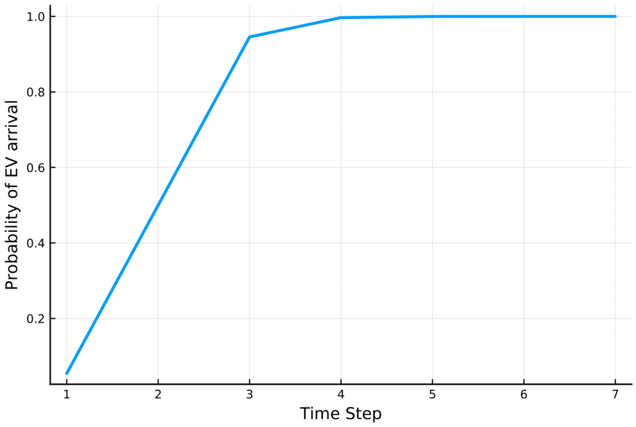

Fig. 5. A probability curve of the EV appearing in any slot at time t given total timescale T

Thus, we use the simple typical probability of the EV arrival (Eq. 3), which based on empirical observations of the driving patterns Metz and Doetsch (2012). Of course, this is a rough approach, however, the main goal of the current study is not to capture more of the real-world dynamics, but to evaluate different MDP solvers for the EV charge optimization problem.

All EVs draw energy from the same grid, which can have a renewable energy level (REL) up to R. Charging a EV will increment its charge level up towards C, the maximum charge level. However charging too many cars at once will reduce the energy level. An ideal policy for this model would discern which cars to charge at every time step, with the goal that by the end of the night (when $t = T$), all cars would be fully or nearly fully charged while keeping the grid's renewable energy mixture level high.

A state-space model is considered to describe the use of the EV. The states are defined by a four component tuple.

(1) A boolean vector of length n describing whether there is a car in each port.
(2) A vector providing the charge level for the EV at that indexed location. Each car can take on discrete charge levels in the range of $[0, C]$. If there is no car present, the charge level reported is zero.
(3) Model car arrival probabilities and renewable level changes as a function of time in the range of $[t, T]$.
(4) The current REL, which can take discrete values in $[0, R]$.

The state space therefore has $|S| = 2^n \cdot (C+1)^n \cdot (R+1)$ possible states. The action to take at each time step is whether to charge each car. The REL is updated based on the number of actions taken and a function modeling how the renewable energy source levels change naturally throughout the day.

$$r_{t+1} = r_t + addPV(t) - addBat(t) - \frac{\sum_n a_i}{n}. \qquad (4)$$

The AddPV(t) and AddBat(t) functions allow to add a level of complexity by modeling how renewable energy changes over time due to external causes. In our default case, we assume there is AddPV(t) or AddBat(t) equal zero, but we also explore how our policy changes when we define these functions.

The reward function is modeled using the current REL r, charge amounts in each EV c, and time t, by

$$R_t(s,a) = \lambda r - \Sigma_{i=1:n} \exp((C - c_i)/C). \qquad (5)$$

The first part of the equation gives reward at each time-step based on the current REL r. This inherently penalizes for dropping the REL. The second part of the equation gives penalties based on the level of charge in each car at the terminal state. Remaining charge is exponentiated to more heavily penalize cars with less charge. The relative weighting of the two components is dictated by λ.

4. EXPERIMENTAL RESULTS

We tested various solvers with a MDP-based environment of the electricity grid including houses with EVs, RELs (wind and PV) and batteries (Fig. 6). Using the POMDPs.jl Julia library Egorov et al. (2017), we evaluated offline MDP solvers (Discrete Value Iteration algorithm, DVI) and online MDP solvers (MCTS) as well as RL-based solvers (Q-learning) to find the optimal policy for EV's flexible charging optimization.

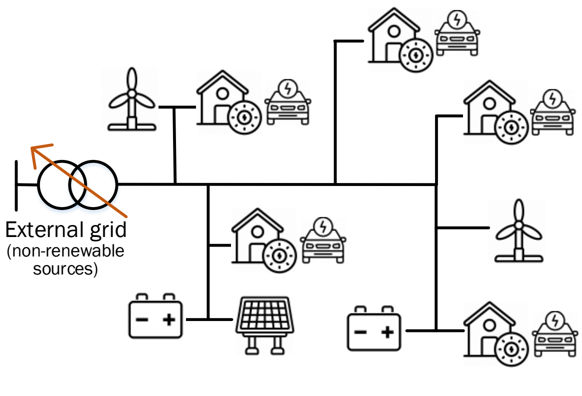

Fig. 6. The example of MDP-based environment of the electricity grid

First of all, we decided to compare the time it takes to find the optimal policy for each solver for a different number of EVs. The results of this comparison are presented in Table

2.It is clearly seen that the offline VDI-based solver does poorly with the scalability of the task. With the increase in the number of EVs, the time spent finding the optimal policy increases significantly.The Q-learning based solver shows significantly better results in terms of calculation time. However, the best results are given by the online MCTS. Therefore, in further experiments we will use the MCTS-based solver.

Table 2. The time to find the optimal policy for MDP solvers with different scalability of the task

Number of EVs	DVI	MCTS	Q-learning
3	145 s	8 s	8 s
4	3108 s	9 s	14 s
5	-	17 s	33 s
6	-	54 s	99 s
7	-	142 s	175 s
8	-	634 s	1011 s
9	-	6985 s	10685 s

Experiments have shown that with decreasing λ, our policy informs us to take the greedy action, and charge every car it can. You can observe for the evolution of policy from more greedy $\lambda = 0.1$ to less greedy $\lambda = 50.0$ policy (Fig. 7–9). The color of each block represents the charge of that EV at that time step, with black indicating a EV is yet to arrive.

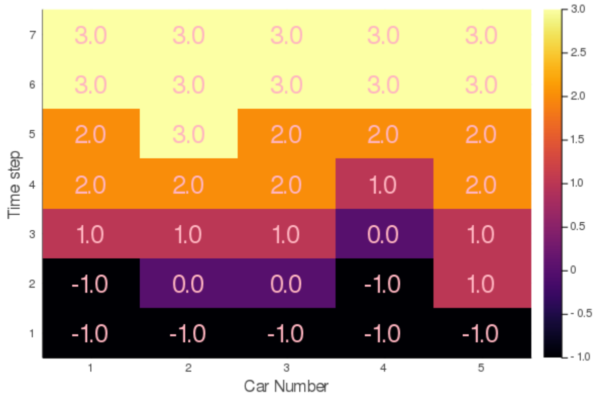

Fig. 7. Charge level 5-car simulation with $\lambda = 0.1$

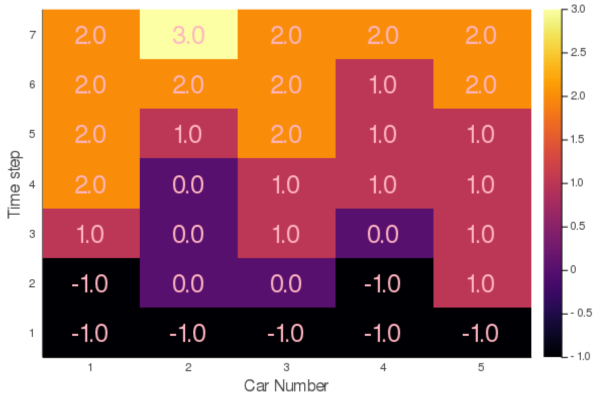

Fig. 8. Charge level 5-car simulation with $\lambda = 10$.

For example, as we can see, $\lambda = 0.1$ policy cares much more than every car is charged by the end of the simulation. However, our MDP-agent try to incrementally charge the cars for most of the time duration in order to keep the REL high. Opposite the $\lambda = 50.0$ policy focuses more on keeping the rel high throughout the simulation. Therefore, the MDP-agent charges only one EV at a time, even if in doing so, the EV will not be fully charged at the end of the simulation.

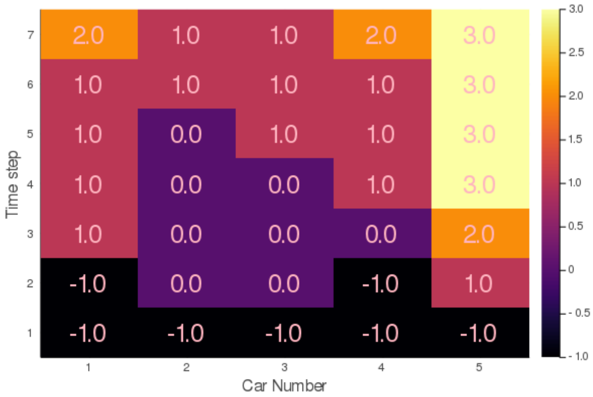

Fig. 9. Charge level 5-car simulation with $\lambda = 50$.

This is because the reward function (5) exponentially penalizes lack of charge, meaning that one car having a small amount of charge is much more costly than having two cars with a moderate amount of charge. According Chandramoul et al. (2018), this mirrors what we would wish to see in real-world implementation of such a tool, since making the decision to not charge one person's car and fully charge another's could lead to consumer distrust of the grid operator.

Next, we observe what happens when we model additional renewable energy being added to the system as it is collected from PV sources. We model to add a level of PV energy for each of the first $T/2$ time steps. This would correspond to starting the simulation when there is daylight still available. It's important to note that even though we use $=10.0$ (less greedy policy) (Fig. 10), our updated policy still informs us to take the greedy action, and charge every car it can during timestep 3, since there would be no reduction in REL in the first few timesteps.

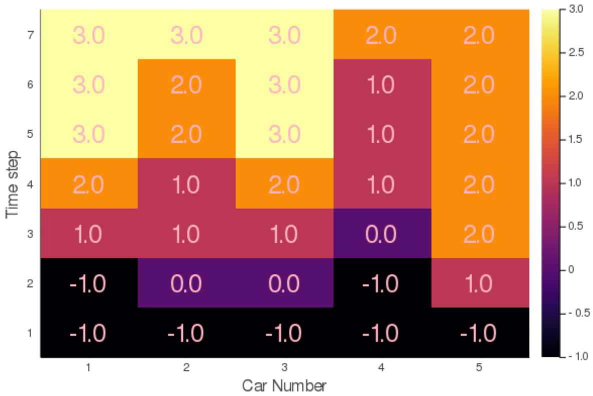

Fig. 10. Charge level 5-car simulation with $\lambda = 10$ with added PV for first $T/2$ time steps

A run with with added batteries for charging on the second $T/2$ time steps indicated that this hypothesis to be true (Fig. 11). We can show a more greedy policy since there would be a reduction in RELs in the second $T/2$ timesteps because we should charge the batteries.

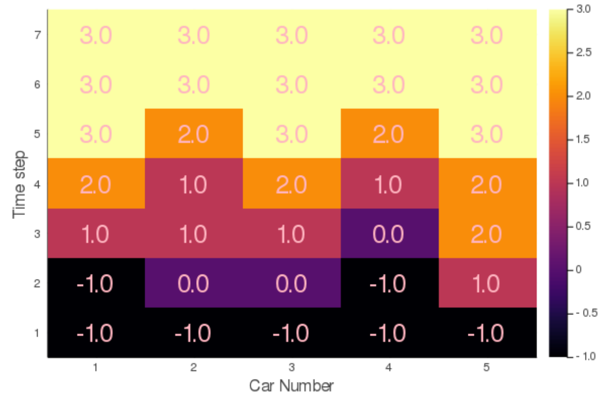

Fig. 11. Charge level 5-car simulation with $\lambda = 10$, added battery charging

5. CONCLUSION

We proposed an MDP-based EV charging process model to find an optimal policy with respect to the number of cars that need to be charged, the length of time available for charging, and different REL dynamics without any prior knowledge of uncertainties. When we added external sources of energy, as expected, the model chose to greedily charge the cars when RELs were high, and become more conservative as these levels dropped (in the case of adding batteries for charging). Due to the curse of dimensionality, standart MDP-based approaches, including RL, fail to solve large-scale problems in real-world scenarios. We have exploited MDPs-based online algorithms such as MCTS to overcome the scalability issues associated with charging of a large number of EVs.

Future work could also utilize other methodologies to capture more of the real-world dynamics and constraints inherent in the problem. For example, cars do not all arrive at the same time, so a future model could utilize a structure more similar to a partially observable MDP to determine an optimal policy despite uncertainty as to each car's arrival time. This could also implement a form of machine learning to determine approximate arrival and departure times for each car over time.

REFERENCES

Brundlinger, T. and et al. (eds.) (2017). *dena-Leitstudie Integrierte Energiewendee.* Deutsche Energie-Agentur GmbH (dena).

Chandramoul, Y., Jamgochian, A., and Jewell, E. (2018). Dope cars: Deciding optimally how to efficiently charge automotives with renewable energy sources.

Egorov, M., Sunberg, Z.N., Balaban, E., Wheeler, T.A., Gupta, J.K., and Kochenderfer, M.J. (2017). POMDPs.jl: A framework for sequential decision making under uncertainty. *Journal of Machine Learning Research*, 18(26), 1–5.

Friedl, G. and et al. (eds.) (2018). *Blackout E-Mobiliat setzt Netzbetreiber unter Druck.* TU Munchen, Munchen.

Hadley, S. and Tsvetkova, A. (2009). Potential impacts of plug-in hybrid electric vehicles on regional power generation. *The Electricity Journal*, 22(10), 56–68.

Jin, C., Mojdehi, M.N., and Ghosh, P. (2012). A methodology to design a stochastic cost efficient der scheduling considering environmental impact. In *2012 International Conference on Smart Grid (SGE)*, 1–6.

J.Maasmann (ed.) (2019). *Die Virtuelle Direktleitung fur den Entfernten Eigenverbrauch durch Elektrofahrzeuge.* Dissertation, Shaker Verlag), Aachen.

Kartal, B., Hernandez-Leal, P., and Taylor, M.E. (2019). Action guidance with MCTS for deep reinforcement learning. *CoRR*, abs/1907.11703. URL `http://arxiv.org/abs/1907.11703`.

Kochenderfer, M. (2015). *Decision Making Under Uncertainty: Theory and Application.* MIT Press, Cambridge.

Li, F. and Du, Y. (2018). From alphago to power system ai: What engineers can learn from solving the most complex board game. *IEEE Power and Energy Magazine*, 16(2), 76–84.

Maasmann, J., Aldejohann, C., Horenkamp, W., Kaliwoda, M., and Rehtanz, C. (2014). Charging optimization due to a fuzzy feedback controlled charging algorithm. In *2014 49th International Universities Power Engineering Conference (UPEC)*, 1–6.

Metz, M.P. and Doetsch, C. (2012). Electric vehicles as flexible loads – a simulation approach using empirical mobility data.

Mukherjee, J.C. and Gupta, A. (2015). A review of charge scheduling of electric vehicles in smart grid. *IEEE Systems Journal*, 9(4), 1541–1553.

Mwasilu, F., Justo, J., Kim, E.K., Do, T.D., and J.-W.Jung (2014). Electric vehicles and smart grid interaction: A review on vehicle to grid and renewable energy sources integration. *Renewable and Sustainable Energy Reviews*, 34, 501–516.

Pregger, T. and et al (eds.) (2012). *Perspektiven von Elektro/Hybridfahrzeugen in einem Versorgungsgebiet mit hohem Anteil dezentraler und erneuerbarer Energiequellen.*

Silver, D., Schrittwieser, J., and K. Simonyan, e.a. (2017). Mastering the game of go without human knowledgei. *Nature*, 550, 354–359.

Sortomme, E. and El-Sharkawi, M.A. (2012). Optimal scheduling of vehicle-to-grid energy and ancillary services. *IEEE Transactions on Smart Grid*, 3(1), 351–359.

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement Learning: An Introduction.* The MIT Press, second edition.

Tang, W., Bi, S., and Zhang, Y.J. (2016). Online charging scheduling algorithms of electric vehicles in smart grid: An overview. *IEEE Communications Magazine*, 54(12), 76–83.

Tomin, N., Kurbatsky, V., and Negnevitsky, M. (2019). Development a partially observable markov decision processes-based intelligent assistant for power grids using monte carlo tree search. In *2 the 10th International Scientific Symposium on Electrical Power Engineering, ELEKTROENERGETIKA 2019*, 389–393.