Multi-Agent Reinforcement Learning for Fast-Timescale Demand Response of Residential Loads

Extended Abstract

Vincent Mai
Mila & Robotics and Embodied AI Lab
Université de Montréal, Canada
vincent.mai@umontreal.ca

Philippe Maisonneuve
Mila & GERAD
Polytechnique Montréal, Canada
philippe.maisonneuve@polymtl.ca

Tianyu Zhang
Mila
Université de Montréal, Canada
tianyu.zhang@mila.quebec

Hadi Nekoei
Mila
Université de Montréal, Canada
nekoeihe@mila.quebec

Liam Paull
Mila & Robotics and Embodied AI Lab
Université de Montréal, Canada
liam.paull@umontreal.ca

Antoine Lesage-Landry
Mila & GERAD
Polytechnique Montréal, Canada
antoine.lesage-landry@polymtl.ca

ABSTRACT

Power grids with high amounts of renewable energy resources must cope with high amplitude, fast timescale variations in power generation. Frequency regulation through demand response has the potential to coordinate temporally flexible loads, such as air conditioners, to counteract these variations. We propose a decentralized agent trained with multi-agent proximal policy optimization with localized communication. We explore two communication frameworks: hand-engineered, or learned through targeted multi-agent communication. The resulting policies perform well and robustly for frequency regulation, and scale seamlessly to arbitrary numbers of houses for constant processing times.

KEYWORDS

Multi-agent reinforcement learning; Demand response; Power systems; Renewable integration; Communication; Coordination

ACM Reference Format:
Vincent Mai, Philippe Maisonneuve, Tianyu Zhang, Hadi Nekoei, Liam Paull, and Antoine Lesage-Landry. 2023. Multi-Agent Reinforcement Learning for Fast-Timescale Demand Response of Residential Loads: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Renewable energy sources such as wind turbines and solar panels are subject to short-term, high-amplitude variations, referred to as intermittency. These creates major challenges when managing the balance between power generation and consumption [12]. At the second timescale, this balancing task is referred to as frequency regulation [3, 23]. The demand response approach [22] aims at adjusting the power demand to meet the supply by coordinating flexible loads temporally [23]. Air conditioners (ACs) are ideal candidates as they represent a significant part of global power consumption. [1, 6]. In this paper, we focus on the task of fast timescale demand response for frequency regulation using *residential ACs*. ACs are *discretely* powered and subject to hardware *dynamic constraints* such as lockout: once turned OFF, they must wait some time before being allowed to turn back ON to protect the compressor. The agents must *cope with uncertainty* in the future regulation signal, be *scalable* to provide enough power flexibility, and *decentralized* with localized communications for implementation considerations. Finally, the decisions must be made *in a few seconds*. These constraints impede the deployment of classical methods. Online Optimization (OO) [13, 14, 29] cannot cope with long-term constraints. Model Predictive Control (MPC) [10, 16, 17, 19, 26] struggles when scaling with the number of agents [7, 10, 15]. We instead tackle this problem with multi-agent reinforcement learning (MARL) to learn decentralized and scalable policies. Our best agents are trained with Multi-Agent Proximal Policy Optimization (MA-PPO) [28] through Centralized Training, Decentralized Execution (CT-DE) [11]. Hand-engineered and learned targeted [8] local communication frameworks are tested – and both outperform the baselines. MARL has been used on longer time scale demand response problems [2, 20, 27] and environments have been developed accordingly [4, 24, 25]. To the best of our knowledge, this is the first usage of MARL for scalable, high frequency demand response using flexible binary loads such as ACs with lockout. Our main contributions are:

- an open source, multi-agent Gym [5] environment[1] simulating the real-world problem of frequency regulation through demand response at the second timescale.
- two decentralized MA-PPO agents[1] with different communication strategies, both outperforming baselines.
- an in-depth analysis of the dynamics, communications, scalability and robustness of the trained agents.

2 PROBLEM FORMULATION

The environment can be described as a decentralized, partially observable Markov decision process (Dec-POMDP). We simulate its dynamics as an aggregation of N houses, each equipped with a single air conditioning (AC) unit controlled by an agent. The thermal dynamics of every house are simulated using a second-order model based on Gridlab-D's Residential module user's guide [9]. The regulation signal, which is the desired aggregated consumption of the ACs, is simulated as the sum of (1) a base signal which covers the

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

[1]The code is available: https://github.com/ALLabMTL/MARL_for_fast_timescale_DR.

Table 1: Performance of the different agents, computed over 10 environment seeds.

	$N_{\text{de}} = 10$		$N_{\text{de}} = 50$		$N_{\text{de}} = 250$		$N_{\text{de}} = 1000$	
Per-agent RMSE	Signal (W)	Max T. (°C)	Signal (W)	Max T. (°C)	Signal (W)	Max T. (°C)	Signal (W)	Max T. (°C)
Greedy	2668 ± 14	0.93	3166 ± 12	1.15	3313 ± 12	1.22	3369 ± 15	1.24
BBC	830 ± 207	0.09	426 ± 63	0.10	318 ± 7	0.10	296 ± 4	0.10
MPC	344 ± 96	0.12	-	-	-	-	-	-
MA-DQN	541 ± 86	0.09	321 ± 24	0.10	246 ± 8	0.11	234 ± 4	0.12
MA-PPO-HE	253 ± 1	0.08	161 ± 8	0.08	127 ± 2	0.11	122 ± 3	0.13
TarMAC-PPO	**247 ± 3**	**0.07**	**158 ± 2**	**0.09**	**115 ± 1**	**0.13**	**101 ± 2**	**0.14**

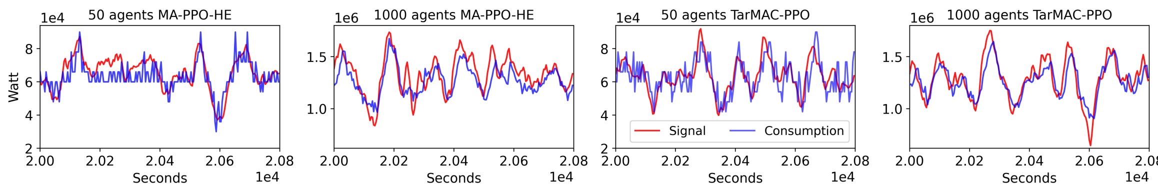

Figure 1: Both MA-PPO policies scale seamlessly in the number of agents: signal and consumption on 800s for $N_{\text{de}} = 50$ and 1000.

needs in energy for each house to maintain an acceptable temperature and (2) a 0-mean Perlin noise simulating the high-frequency variations. Agents observe the indoor and outdoor temperatures, the state of the AC and its lockout time, and the current per-agent signal and total consumption of the aggregation. They act by turning the AC ON or OFF, constrained by a lockout to protect the compressor. In addition, agents can communicate. To keep the implementation decentralized and flexible, each agent can only exchange information with a number N_c of neighbours. The reward for each agent is the weighted sum of squared penalties due to (1) the house's air temperature being different from the target, which is unique to the agent, and to (2) the consumption signal tracking, which is common across all agents.

3 AGENTS

Two types of MARL agents were trained on this environment using the CT-DE paradigm. MA-DQN, based on the Deep Q-Network [18] algorithm, and MA-PPO [28], based on PPO [21]. We implemented two variations for communications between agents: in the hand-engineered (HE) version, applied to MA-DQN and MA-PPO, the agents send a predetermined part of their observations as the messages to their neighbours. These messages are concatenated with the receiver's own observations as the input to the policy, fixing the number of houses an agent communicates with. We also implemented a MA-PPO version of TarMAC [8], where message contents are learned and the received messages are aggregated based on an attention mechanism. This allows a more flexibility as per the number of houses each agent communicates with. We compare their performance with a bang-bang controller (BBC) for temperature and classical and centralized baselines such as a greedy myopic knapsack solver and a model predictive controller (MPC).

4 RESULTS AND ANALYSIS

We deploy the agents on a benchmark environment with N_{de} houses on trajectories of 43200 steps of 4 seconds. We measure the per-agent root mean square error (RMSE) between the regulation signal s_t and aggregated power consumption P_t and the temperature RMSEs of the maximal temperature error of the aggregation. Table 1 shows the performance of different agents in environments with and without lockout with N_{de} of 10, 50, 250 and 1000 houses. BBC tracks the temperature but not the signal. Greedy myopic fails: it does not plan for the lockout and runs out of available agents. MPC gives good results for 10 agents, but could not be run on $N_{\text{de}} = 50$ for computing time reasons. DQN controls the temperature well but is only slightly better than BBC on the signal. The PPO agents show significantly better performance. Both scale gracefully with the number of agents, but TarMAC-PPO outperforms MA-PPO-HE at high N_{de}. Figure 1 shows the consumption and signal over 800 seconds for both agents deployed over $N_d = 50$ and 1000 over 800 seconds. For $N_d = 50$, they do not perfectly match the signal. However, the same agents do better on 1000 houses. Indeed, as the environment is homogeneous, local errors average out when scaling. We remarked that MA-PPO-HE learned cyclic coordination patterns through their fixed message structure, while such patterns were absent from TarMAC-PPO. Further experiments showed that the best performing PPO agents were trained on environments with $N_{\text{tr}} = 10$ houses, as training with more agents makes credit assignment harder. We also observed that communicating with only 9 neighbours often leads to the best performance. We further show that TarMAC-PPO is robust to faulty communications, heterogeneous houses and ACs, and environmental shifts.

5 CONCLUSION

In this work, we tackle the problem of high-frequency regulation with demand response by controlling discrete and dynamically constrained residential loads equipped with air conditioners with a decentralized, real-time agent trained by MA-PPO with hand-engineered messages or learned targeted communication. The policies perform significantly better than baselines, scale seamlessly to large numbers of houses, and are robust to most disturbances. Our results show that MARL can be used successfully to solve some of the complex multi-agent problems induced by the integration of renewable energy in electrical power grids.

ACKNOWLEDGMENTS

This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Institute for Data Valorization (IVADO).

REFERENCES

[1] International Energy Agency. 2018. *The Future of Cooling.* https://www.iea.org/reports/the-future-of-cooling

[2] Sally Aladdin, Samah El-Tantawy, Mostafa M. Fouda, and Adly S. Tag Eldien. 2020. MARLA-SG: Multi-Agent Reinforcement Learning Algorithm for Efficient Demand Response in Smart Grid. *IEEE Access* 8 (2020), 210626–210639. https://doi.org/10.1109/ACCESS.2020.3038863

[3] Hassan Bevrani, Arindam Ghosh, and Gerard Ledwich. 2010. Renewable energy sources and frequency regulation: survey and new perspectives. *IET Renewable Power Generation* 4, 5 (2010), 438–457.

[4] David Biagioni, Xiangyu Zhang, Dylan Wald, Deepthi Vaidhynathan, Rohit Chintala, Jennifer King, and Ahmed S. Zamzam. 2021. PowerGridworld: A Framework for Multi-Agent Reinforcement Learning in Power Systems. https://doi.org/10.48550/ARXIV.2111.05969

[5] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. arXiv:arXiv:1606.01540

[6] Duncan S Callaway. 2009. Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Conversion and Management* 50, 5 (2009), 1389–1400.

[7] Bingqing Chen, Jonathan Francis, Marco Pritoni, Soummya Kar, and Mario Bergés. 2020. COHORT: Coordination of Heterogeneous Thermostatically Controlled Loads for Demand Flexibility. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.* 31–40. https://doi.org/10.1145/3408308.3427980 arXiv:2010.03659 [cs, eess].

[8] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Mike Rabbat, and Joelle Pineau. 2019. TarMAC: Targeted Multi-Agent Communication. In *Proceedings of the 36th International Conference on Machine Learning*. PMLR, 1538–1546. https://proceedings.mlr.press/v97/das19a.html

[9] Betelle Memorial Institute. [n.d.]. GridLAB-D Wiki. http://gridlab-d.shoutwiki.com/wiki/Main_Page http://gridlab-d.shoutwiki.com/wiki/Main_Page (Accessed on: Sept 15, 2022).

[10] Jin, Mohammed Olama, Teja Kuruganti, James Nutaro, Christopher Winstead, Yaosuo Xue, and Alexander Melin. 2018. Model Predictive Control of Building On/Off HVAC Systems to Compensate Fluctuations in Solar Power Generation. In *2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. 1–5. https://doi.org/10.1109/PEDG.2018.8447840

[11] Landon Kraemer and Bikramjit Banerjee. 2016. Multi-agent reinforcement learning as a rehearsal for decentralized planning. *Neurocomputing* 190 (May 2016), 82–94. https://doi.org/10.1016/j.neucom.2016.01.031

[12] Prabha Kundur. 2007. Power system stability. *Power system stability and control* (2007), 7–1.

[13] Antoine Lesage-Landry and Joshua A Taylor. 2018. Setpoint tracking with partially observed loads. *IEEE Transactions on Power Systems* 33, 5 (2018), 5615–5627.

[14] Antoine Lesage-Landry, Joshua A Taylor, and Duncan S Callaway. 2021. Online convex optimization with binary constraints. *IEEE Trans. Automat. Control* 66, 12 (2021), 6164–6170.

[15] M Liu and Y Shi. 2015. Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services. *IEEE Trans. on Power Systems* 31, 3 (2015), 1963–1971.

[16] Mehdi Maasoumy, Borhan M Sanandaji, Alberto Sangiovanni-Vincentelli, and Kameshwar Poolla. 2014. Model predictive control of regulation services from commercial buildings to the smart grid. In *2014 American Control Conference*. IEEE, 2226–2233.

[17] J L Mathieu, S Koch, and D S Callaway. 2012. State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Trans. on Power Systems* 28, 1 (2012), 430–440.

[18] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. 2015. Human-level control through deep reinforcement learning. *Nature* 518, 75407540 (Feb 2015), 529–533. https://doi.org/10.1038/nature14236

[19] Mohammed M. Olama, Teja Kuruganti, James Nutaro, and Jin Dong. 2018. Coordination and Control of Building HVAC Systems to Provide Frequency Regulation to the Electric Grid. *Energies* 11, 7 (2018). https://doi.org/10.3390/en11071852

[20] Martin Roesch, Christian Linder, Roland Zimmermann, Andreas Rudolf, Andrea Hohmann, and Gunther Reinhart. 2020. Smart Grid for Industry Using Multi-Agent Reinforcement Learning. *Applied Sciences* 10, 19 (2020). https://doi.org/10.3390/app10196900

[21] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. (2017). https://doi.org/10.1007/s00038-010-0125-8

[22] Pierluigi Siano. 2014. Demand response and smart grids—A survey. *Renewable and sustainable energy reviews* 30 (2014), 461–478.

[23] Josh A Taylor, Sairaj V Dhople, and Duncan S Callaway. 2016. Power systems without fuel. *Renewable and Sustainable Energy Reviews* 57 (2016), 1322–1336.

[24] Jose R. Vazquez-Canteli, Sourav Dey, Gregor Henze, and Zoltan Nagy. 2020. CityLearn: Standardizing Research in Multi-Agent Reinforcement Learning for Demand Response and Urban Energy Management. (Dec 2020). https://doi.org/10.48550/arXiv.2012.10504 arXiv:2012.10504 [cs].

[25] Jose R. Vazquez-Canteli, Gregor Henze, and Zoltan Nagy. 2020. MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (Virtual Event, Japan) *(BuildSys '20)*. Association for Computing Machinery, New York, NY, USA, 170–179. https://doi.org/10.1145/3408308.3427604

[26] Xiaoyu Wu, Jinghan He, Yin Xu, Jian Lu, Ning Lu, and Xiaojun Wang. 2018. Hierarchical Control of Residential HVAC Units for Primary Frequency Regulation. *IEEE Transactions on Smart Grid* 9, 4 (2018), 3844–3856. https://doi.org/10.1109/TSG.2017.2766880

[27] Yaodong Yang, Jianye Hao, Yan Zheng, Xiaotian Hao, and Bofeng Fu. 2019. Large-Scale Home Energy Management Using Entropy-Based Collective Multiagent Reinforcement Learning Framework. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2285–2287.

[28] Chao Yu, Akash Velu, Eugene Vinitsky, Yu Wang, Alexandre Bayen, and Yi Wu. 2021. The Surprising Effectiveness of PPO in Cooperative, Multi-Agent Games. (Jul 2021). arXiv:2103.01955 [cs].

[29] X Zhou, E Dall'Anese, and L Chen. 2019. Online stochastic optimization of networked distributed energy resources. *IEEE Trans. on Automatic Control* 65, 6 (2019), 2387–2401.