Multi-agent reinforcement learning for renewable integration in the electric power grid

Vincent Mai[1], **Tianyu Zhang**[1] **& Antoine Lesage-Landry**[2]
[1]Mila, [2]Polytechnique Montréal & GERAD

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021

Tuesday December 14th, 2021

Université de Montréal

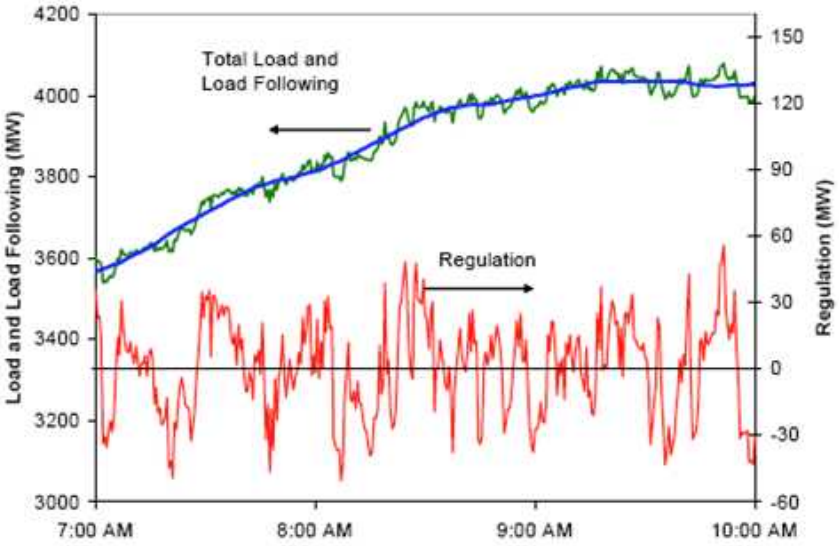
Mila

GERAD
GROUP FOR RESEARCH IN DECISION ANALYSIS

POLYTECHNIQUE MONTRÉAL
TECHNOLOGICAL UNIVERSITY

Renewable integration & challenges

(a) Uncertain, intermittent, fast-ramping renewables *Source: NRDC/Vanja Terzic/iStock*



(b) Sustained need for balancing generation & demand *Source: Oak Ridge National Lab*

Figure 1: Paradigm shift in electric grid operations

- Frequency regulation: load balancing on short timescale;
- Stability of the grid;
- **Demand response of thermostatically controlled loads**
  - modulate power consumption of air conditioners/heating to mitigate renewable intermittency.
- Alternatives: fuel-burning power plant or expensive batteries.

Multi-agent reinforcement learning-based demand response

Core challenges in frequency regulation with thermostatic loads:

1. Uncertainty:
   - limited measurements or feedback;
   - exogenous factors to the power systems, e.g., weather.
2. Real-time decision-making:
   - requires power adjustment every few seconds.
3. Dynamic load constraints:
   - temperature preferences & lock out of air conditioners.
4. Scalability to & cooperativity in multi-agent settings:
   - coordination of many loads $\rightarrow$ consumption meets objective.

Multi-agent PPO to coordinate the power consumption of residential households and provide frequency regulation.

1. policy $\rightarrow$ adaptive;
2. policy $\rightarrow$ only requires evaluation to obtain decision;
3. Bellman equation in RL $\rightarrow$ dynamic environments;
4. centralized training & decentralized execution $\rightarrow$ scalability.

Problem formulation

The demand response process is set as an POMDP. At time t:

1. Each TCL i observes its state $\mathbf{X}_{i,t}$:
   - current, target and outdoors temperatures
   - AC power and lockdown status
   - regulation signal and aggregated consumption
   - time and date
   - thermal characteristics of the TCL
   - communication from neighbours
2. Each TCL takes an action $u_{t,i}$: turn the TCL ON or OFF.
3. Each TCL receives a reward $r_{t,i}$ based on:
   - the tracking of the temperature target (different for each TCL)
   - the tracking of the regulation signal by the aggregated power consumption (same for all TCLs)

Multi-agent environment

The multi-agent environment is based on OpenAI Gym and Ray.

- multiple TCLs: from 1 to +1000
- second order thermal model inspired by GridLAB-D
- outdoors temperature using real data
- noisy regulation signal: first artificial, then real data.

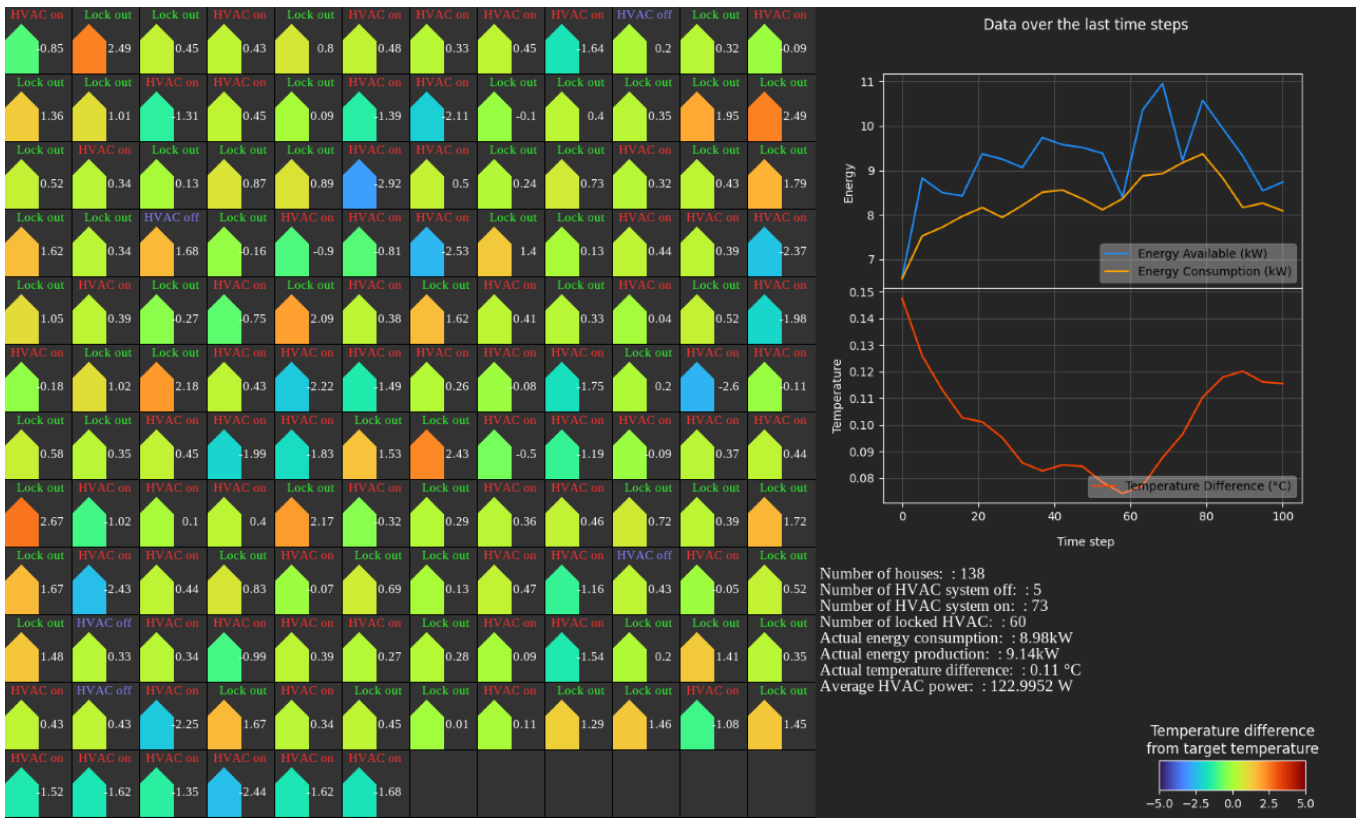


Figure 2: Rendering of the environment.

Multi-agent reinforcement learning

- Proximal Policy Optimization (PPO)
- Off-policy policy gradient
- Stability and performance improvements by adding constraints between the on-policy and the off-policy actors.

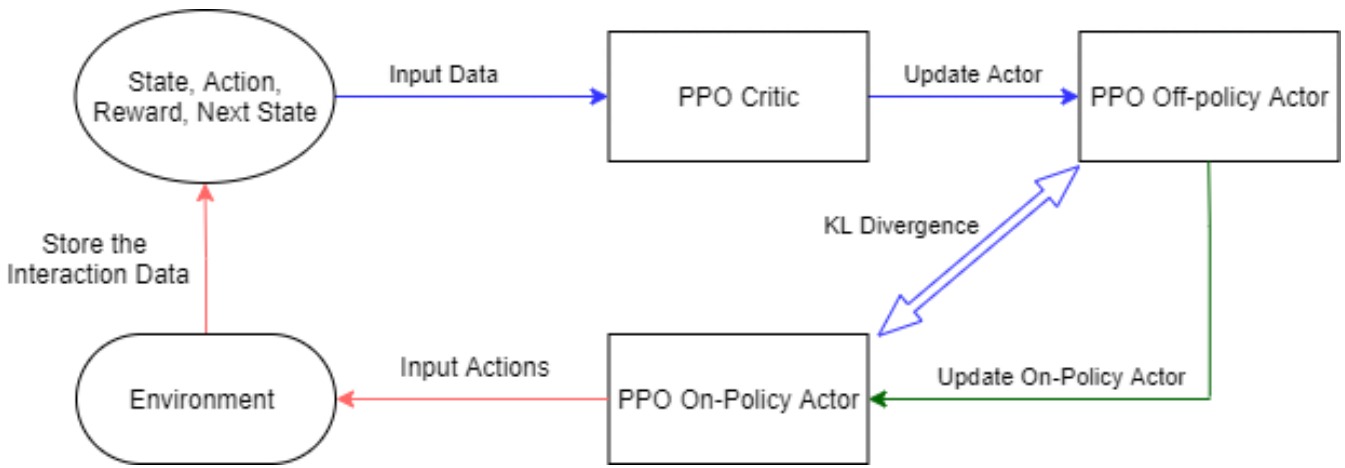


Figure 3: 2-stage PPO training scheme: 1) data gathering and 2) agent training.

Multi-agent reinforcement learning

- Parameter sharing with centralized training and decentralized execution
- Hindsight experience replay: a multi-goal, single agent problem
- Curriculum Learning: design sub-goals for the agents to boost training

Conclusion & collaboration

To tackle the uncertainty and dynamic aspects of thermostatic loads in DR, we formulate a multi-agent reinforcement learning approach for frequency regulation.

Looking for collaborators for:

- Building & smart grid modelling expertise;
- Building simulators (general, AC/heating or water heaters);

Future work

- Safety: network-aware demand response;
- Simulation to real implementation (sim-to-real);
- Extension to other flexible loads (e.g. EVs.);
- Experimental case study.

Vincent Mai
vincent.mai@umontreal.ca

Tianyu Zhang
tianyu.zhang@mila.quebec

Antoine Lesage-Landry
antoine.lesage-landry@polymtl.ca

This work was funded by the Institute for Data Valorization (IVADO), by the National Sciences and Engineering Research Council of Canada, as well as Microsoft and Samsung.

Appendix