Decentralized Partially Observable Markov Decision Process for Power Grid Management

Eva Boguslawski, Alessandro Leite, Marc Schoenauer, Matthieu Dussartre, Benjamin Donnot

Decentralized Partially Observable Markov Decision Process for Power Grid Management

Eva Boguslawski, Alessandro Leite, Marc Schoenauer, Matthieu Dussartre, Benjamin Donnot
RTE, TAU, INRIA, Université Paris-Saclay
eva.boguslawski@rte-france.com

1. Context

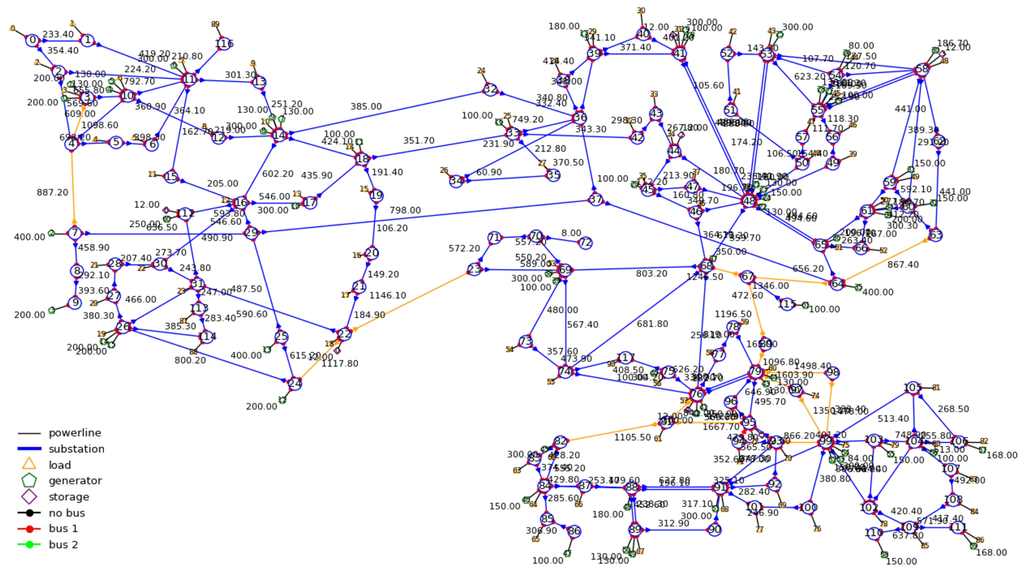
Renewable energy sources are forcing power grid operators to review their grid management strategies

Réseau de Transport d'Électricité (RTE) is developing new adaptive zonal automatons

- **Each automaton monitors a zone of the power grid** (up to 40 lines) thanks to an optimization algorithm.
- An automaton can act on the power grid configuration and on the electricity production.
- **Each automaton can receive a target path or additional information** from operators to decide the appropriate actions to be done. Examples of target path include (un)desirable configurations in future hours or information about other zones.

2. Objective

Design a decision support assistant for the supervision of zonal automatons able to **recommend relevant target paths/information for automatons** based on the power grid's configuration and constraints

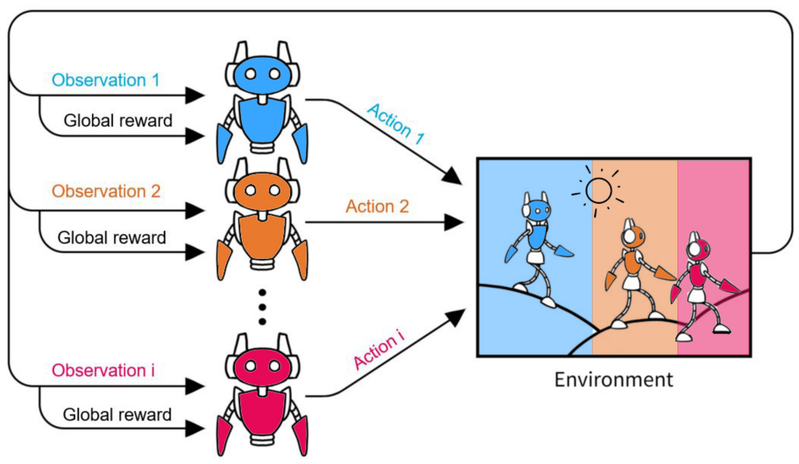


Example of a simulated power grid

3. Collaborative Multi-Agent Reinforcement Learning

A **decentralized approach can improve the scalability** compared to traditional reinforcement learning methods.

The supervision of automatons can be modeled as a **Decentralized Partially Observable Markov Decision Process** (Dec-POMDP) [1].



Decentralized multi-agent reinforcement learning

For example, each agent can be associated with a sub-group of automatons and their zones. At each timestep t:

1. Agent i gets an observation of the i^{th} sub-group of zones
2. Agent i chooses an action according to its observation. An action means here, transmitting a chosen target paths to the i^{th} sub-group of automatons.
3. Actions of agents are combined and executed on the environment
4. A global reward representing the safety of the hole grid is returned

Advantages and limits:

- A decentralized strategy would allow to **reduce the observation and action space and enable inference in parallel**.
- Literature lacks decentralized policies under stochastic conditions

4. How to emulate a zonal automaton?

To obtain **reasonable computation times** during training, we need a trade-off between speed and realism.

We will try to emulate an automaton with a RL agent able to:

1. Operate the grid
2. Follow a target setpoint

We designed an agent able to operate a 14-nodes power grid by controlling storage power

Performance of our agent on a 14-nodes simulated power grid

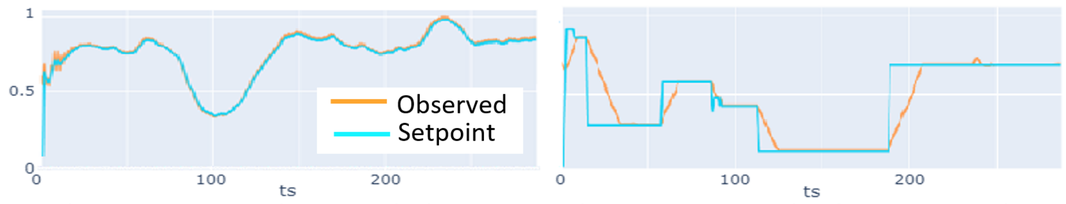
	DoNothing Agent	RL Agent
Average number of survived time steps per episode	206/288	**279/288**

We designed an agent able to follow a storage charge setpoint

We got interesting results about:

- how to penalize deviations from the setpoint.
- the impact of hyperparameters.
- the ability to **generalize to other shapes of setpoint** which is great as depicted in the figure below.

Normalized storage charge



Charge setpoint with a smooth shape | Charge setpoint with a function step shape

Comparison of observed charge with a smooth and a step function setpoint

Open questions

1. How to model network configuration under different levels of uncertainties and agents' objectives?
2. How do Dec-POMDP methods perform under transient network configuration and with hundreds of nodes?
3. How to combine model-based MARL and Dec-POMDP to have a performance independent of reward function and avoid reward hacking?

References

[1] Karl Johan Åström. Optimal control of markov processes with incomplete state information. *Journal of mathematical analysis and applications*, 10(1):174–205, 1965.