AI-based control of smart grids: analysis of the vulnerabilities of an AI-based algorithm used to control an electric vehicles recharging infrastructure

(Internship description)

Keywords

Machine Learning, Data Poisoning (Cyber) Attacks, Multi-Armed Bandits, Adaptive Multi-Agents, Smart Charging,

Supervisors and associated researchers

Three main supervisors:

- Anne Blavette, CNRS, SATIE/IETR, ENS Rennes, anne.blavette@ens-rennes.fr,
- Raphaël Féraud, Orange Labs, Lannion, raphael.feraud@orange.com,
- Michel Hurfin, Inria center of the University of Rennes, PIRAT team, michel.hurfin@inria.fr

Four other associated researchers :

- Hamid Ben Ahmed, SATIE, ENS Rennes, benahmed@ens-rennes.fr,
- Gilles Guette, University of Rennes, PIRAT team, gilles.guette@univ-rennes.fr,
- Yufei Han, Inria center of the University of Rennes, PIRAT Team PIRAT, yufei.han@inria.fr,
- Guy Camilleri, IRIT, Université Paul Sabatier, Toulouse, guy.camilleri@irit.fr

Location

Project team PIRAT at CentraleSupélec, Rennes
(frequent visits are to be expected in the premises of ENS Rennes and Inria Rennes, as well as a short research stay at Orange)

Duration

5 or 6 months, starting date between January and May 2024.

Context

Electric mobility is currently booming and this growth in demand is accompanied by a progressive deployment of recharging infrastructures and associated services which require the design of new control methods. For instance, for large-scale fleets of light EVs, a centralised control is impossible due to privacy issues and a prohibiting computing effort for optimising the charging (and even discharging) of these EVs. Hence, decentralized control methods (some of which based on machine learning) are envisaged.

In this general context, a collaboration between members of the SATIE lab, the IRIT institute and Orange has already led to the design of a machine learning-based algorithmic solution to the smart charging problem of large-scale EV fleets [4]. The solution is based on combinatorial multi-armed bandits decentralised by an adaptive multi-agent system. Reinforcement learning algorithms (and more specifically, the use of multi-armed bandits [1]) allow agents, which control the use of charging stations, to converge quickly towards a sub-optimal, but reasonably good (dis)charging profile.

The solution proved to be scalable and able to satisfy the constraints (congestion, EV mobility needs) for a value of the objective function relatively close to the optimal solution (as simulated in small-scale studies). Yet, like any IT infrastructure, the smart charging infrastructure (e.g. charging stations, electric vehicles, etc.) can be the target of malicious attackers that may affect (in the worst case scenario) the whole grid system. Indeed simultaneous cyberattacks may lead to serious consequences, such as blackouts.

Objectives

The objectives of this internship are:

1. to identify some possible attacks against the proposed AI-based algorithm
2. to develop realistic demonstrators of these attacks,
3. to propose appropriate metrics to measure the power of the attacker and the impact of their malicious actions on the electrical network.

More precisely, the work will be divided into two mandatory steps (plus one optional):

First step: study of the state of the art on attacks targeting :

a) electric vehicle charging infrastructures in general
b) AI-based algorithms in general (either during the learning phase like for example poisoning attacks [3] or during the exploitation of the model like for example membership inference attack [2])

Second step: focus on the algorithm proposed to control recharging stations
a) what attacks are applicable to this algorithm (either via the AI tools used or independently of them) ?
b) how the algorithm behaves if it interacts with already corrupted entities (vehicles, charging stations, agents, etc.) and/or in an environment degraded due to a cyber-attack ?
c) what are the impacts and possible consequences of these attacks if the algorithm was used in a real infrastructure ?
d) implementations of some of these attacks will be done with the aim to quantify the observable deviations and to propose specific metrics to evaluate the impact (periods of system unavailability, violations of constraints expressed by the grid operator, increased convergence time, etc.)

Third step (optional): countermeasures (existing in the literature or not) and possible changes to be made to the original algorithm

This preliminary work is envisaged to be followed by a PhD thesis starting in September 2024.

Required skills

We expect the candidate to have good understanding of Machine Learning, especially reinforcement learning. A good knowledge in optimization in general and in security would also be appreciated.

The beginning of the internship is devoted to a bibliographic study and an understanding of the targeted algorithm. Thus, if necessary, the trainee will have the opportunity to supplement his/her initial knowledge with the help of the supervisors.

An interest in multi-disciplinary topics would be a strong plus, as the candidate will be required to learn about power systems and cybersecurity. Training will be provided by the SATIE and IRISA labs.

Contacts

We invite candidates to forward their CV to all the supervisors and associated researchers.
Please contact any of us for more details on this internship project.

References

[1] Raphaël Féraud. On the relevance of bandit algorithms in digital world, HDR, University Paris-Saclay, France, 2023.

[2] Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, Philip S. Yu, and Xuyun Zhang. Membership inference attacks on machine learning: A survey. *ACM Comput. Surv.*, 54(11s), sep 2022.

[3] Zhiyi Tian, Lei Cui, Jie Liang, and Shui Yu. A comprehensive survey on poisoning attacks and countermeasures in machine learning. *ACM Comput. Surv.*, 55(8), dec 2022.

[4] Sharyal Zafar, Raphaël Féraud, Anne Blavette, Guy Camilleri, and Hamid Ben Ahmed. Decentralized smart charging of large-scale EVs using adaptive multi-agent multi-armed bandits. In *Proc. CIRED, Rome, Italy*, 2023.