IoT Sensor Gym: Training Autonomous IoT Devices with Deep Reinforcement Learning

Abdulmajid Murad
Norwegian University of Science and Technology, NTNU
Trondheim, Norway
abdulmajid.a.murad@ntnu.no

Frank Alexander Kraemer
Norwegian University of Science and Technology, NTNU
Trondheim, Norway
kraemer@ntnu.no

Kerstin Bach
Norwegian University of Science and Technology, NTNU
Trondheim, Norway
kerstin.bach@ntnu.no

Gavin Taylor
United States Naval Academy, USNA
Annapolis, USA
taylor@usna.edu

ABSTRACT

We describe *IoT Sensor Gym*, a framework to train the behavior of constrained IoT devices using deep reinforcement learning. We focus on the main architectural choices to align problems from the IoT domain with cutting-edge reinforcement learning algorithms and exemplify our results with the autonomous control of a solar-powered IoT device.

CCS CONCEPTS

• **Computer systems organization** → **Sensor networks**; • **Theory of computation** → **Reinforcement learning**.

IoT2019, October 22-25, 2019, Bilbao, Spain

Table 1: Analogy between gaming concepts and resource-constrained IoT devices.

Games	→	IoT
Player 	→	IoT Device Agent
World 	→	Weather and other relevant conditions.
Reward 	→	Data measurements
Control 	→	Duty cycle, sensing frequency,...

The source code is available at

https://github.com/Abdulmajid-Murad/IoT-Sensor-Gym.

The code provides the functionality to train agents using OpenAI baseline implementations of reinforcement learning algorithms [4], including PPO [9]. Further, it provides functionality to evaluate the trained agents, as well as saving and restoring them.

KEYWORDS

Deep Reinforcement Learning; Internet of Things; IoT; Embedded Systems; Energy Management

ACM Reference Format:
Abdulmajid Murad, Frank Alexander Kraemer, Kerstin Bach, and Gavin Taylor. 2019. IoT Sensor Gym: Training Autonomous IoT Devices with Deep Reinforcement Learning. In *Proceedings of International Conference on Internet of Things (IoT2019).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3365871.3365911

INTRODUCTION

With constrained resources and the need for scalable solutions, individual and autonomous control through behavior learning is crucial for IoT devices so that they can optimally act in dynamic and non-stationary environments. However, current IoT solutions often rely on manual, static configurations or fined-tuned algorithms that fail to suit all nodes of large-scale IoT systems. Fortunately, reinforcement learning emerged as a promising approach to automate IoT control, and it has already been used in various IoT applications [6] to solve simpler problems. Lately, deep reinforcement learning (DRL) proved to be effective in also solving harder problems due to its ability to model continuous observation spaces, as well as improved function approximation, which lead to its application in many fields, from continuous control tasks [5] to Atari games [1]. In this work, we explore how to apply DRL to the domain of IoT, inspired by the analogy between gaming and IoT, illustrated in Table 1.

THE SENSOR-GYM DESIGN

We built the IoT Sensor Gym as an extension to the OpenAI Gym framework [3]. Sensor Gym provides an environment specific to constrained IoT devices, with an emphasis on their energy budget. In DRL, an agent is built using a neural network and learns a policy through experience by interacting with the environment. At each time step, the agent observes the environment's state $\mathbf{s_t}$ and selects an action $\mathbf{a_t}$. Depending on the state and the selected action, the agent receives a reward value $\mathbf{R_t}$. This reward is used for learning via various learning algorithms. One recent algorithm is Proximal Policy Optimization (PPO) [9]. It is a policy-gradient method able to solve also harder learning problems with little problem-specific engineering and is capable of using continuous values for states and actions.

Figure 1 illustrates the architecture of the IoT Sensor Gym framework and the process of training and deploying RL agents. IoT devices are simulated using a variety of models which can be combined and configured to match various use cases. The energy-buffer model simulates an energy buffer of a maximum capacity $\mathbf{B}_{max}$, and provides the level of energy at each time-step $\mathbf{B}_t$. The energy-harvesting model calculates the harvested energy $\mathbf{Eh}_t$, using a selected solar panel and location-dependent solar radiation data. The energy-consumption model simulates a load that consumes energy at each time step $\mathbf{Ec}_t$ according to a selected duty cycle $\mathbf{D}_t$ (i.e., $\mathbf{Ec}_t \propto \mathbf{D}_t$). Accordingly, the next level of the energy

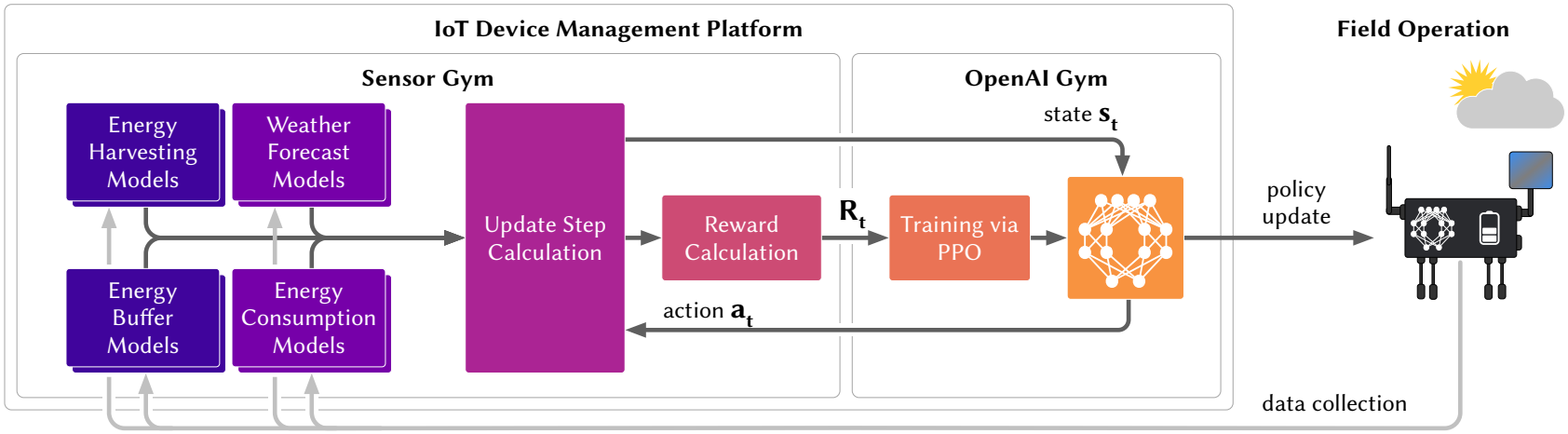

Figure 1: Architecture of the IoT Sensor Gym as a framework to train the behavior of constrained IoT devices using deep reinforcement learning

buffer is given by $B_{t+1} = \min(B_t + Eh_t - Ec_t, B_{max})$. The weather forecast model provides general information about the expected weather in terms of estimated solar energy, and it can be acquired from external sources or prediction algorithms as in presented in [7]. More specific models can be added to suite also other settings.

The *Update Step* calculation acts as a bridge between Sensor Gym and an external DRL agent to run one step of the environment's dynamics. The interaction between Sensor Gym and an RL agent starts with creating a simulated IoT environment and initializing it with the default values. It also defines the boundaries of the environment, such as allowed actions (action space), the environment's possible states (state space), as well as sampling and returning the first state. Then, the interaction proceeds through *Update Step* by running one epoch of the environment's dynamics after receiving an action from the agent. The *training* of agents occurs off-board on a server, for instance as a part of the IoT device management [2]. The trained agents are then deployed to real IoT devices and updated regularly. The agent's policies can usually be approximated with neural networks that require acceptable computational effort and memory footprint, so that they can also be deployed in modern, energy-efficient IoT devices. The models required for the simulation in the IoT Sensor Gym can be updated with data observed by the sensor device. This requires corresponding instrumentation, for instance by measuring its energy intake and consumption, indicated by the feedback in Figure 1.

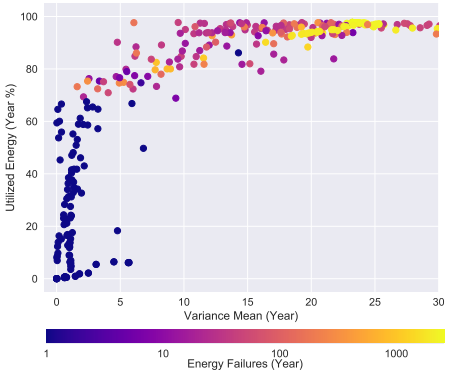

Figure 2: Performance results of over 300 agents. The x-axis corresponds to the mean variance of the duty cycle, the y-axis to the yearly utilized energy. Each dot represents an agent, and the color indicates the number of times an agent has emptied its energy buffer, i.e., failed.

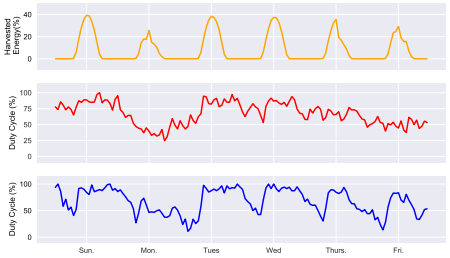

Figure 3: Solar energy intake and resulting duty cycle selection of two agents. The red agent receives more penalty for duty cycle variance than the blue one and is hence smoother, but utilizes slightly less energy.

USE CASE: DUTY-CYCLE OPTIMIZATION UNDER UNCERTAIN ENERGY-HARVESTING

In this use case, a solar-powered IoT device with relatively small energy buffer should maximize its duty cycle, i.e., utilize as much of the incoming solar energy as possible, but without failing by depleting its buffer. At the same time, the duty cycle should have a low variance to ensure steady data collection. For that, we designed a corresponding reward function that reflects these application goals [8], using hyperparameters to balance between them. Figure 2 shows the performance of more than 300 different agents over a whole year, using different hyperparameters. Figure 3 shows the resulting duty cycles of two selected agents over six days. The first day is sunny, so the agents utilize this energy by selecting a high duty-cycle. Then, they reduce their duty cycles, since the weather forecast indicates less energy. The agents have different hyperparameters that model the tradeoff between energy utilization and low variance, hence illustrating the tradeoffs that the approach enables.

CONCLUSION

The IoT Sensor Gym provides the basic mapping of behavior control of constrained IoT devices using energy harvesting to cutting-edge reinforcement learning algorithms. This opens for the control IoT devices by autonomous agents that are able to make complex decisions and learn in non-stationary environments also for more difficult problems.

REFERENCES

[1] Marc G. Bellemare, Yavar Naddaf, Joel Veness, and Michael Bowling. 2013. The Arcade Learning Environment: An Evaluation Platform for General Agents. *Journal of Artificial Intelligence Research* 47, 1 (May 2013), 253–279.

[2] Anders Eivind Braten and Frank Alexander Kraemer. 2018. Towards Cognitive IoT: Autonomous Prediction Model Selection for Solar-Powered Nodes. In *Int. Congress on Internet of Things (ICIOT)*. IEEE, https://doi.org/10.1109/ICIOT.2018.00023

[3] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. *arXiv.org* abs/1606.01540 (2016), 4. arXiv:arXiv:1606.01540

[4] Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, Yuhuai Wu, and Peter Zhokhov. 2017. OpenAI Baselines. https://github.com/openai/baselines.

[5] Yan Duan, Xi Chen, Rein Houthooft, John Schulman, and Pieter Abbeel. 2016. Benchmarking Deep Reinforcement Learning for Continuous Control. In *33rd Int. Conf. on Machine Learning (ICML'16)*. JMLR.org, 1329–1338.

[6] Francesco Fraternali, Bharathan Balaji, and Rajesh Gupta. 2018. Scaling Configuration of Energy Harvesting Sensors With Reinforcement Learning. In *6th Int. Ws. on Energy Harvesting & Energy-Neutral Sensing Systems*. ACM, 7–13. https://doi.org/10.1145/3279755.3279760

[7] Frank Alexander Kraemer, Doreid Ammar, Anders Eivind Braten, Nattachart Tamkittikhun, and David Palma. 2017. Solar Energy Prediction for Constrained IoT Nodes Based on Public Weather Forecasts. In *IoT '17 Proceedings of the Seventh International Conference on the Internet of Things*. ACM, 1–8. https://doi.org/10.1145/3131542.3131544

[8] Abdulmajid Murad, Frank Alexander Kraemer, Kerstin Bach, and Gavin Taylor. 2019. Autonomous Management of Energy-Harvesting IoT Nodes Using Deep Reinforcement Learning. In SASO 2019, IEEE, https://doi.org/10.1109/SASO.2019.00015

[9] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *ArXiv* abs/1707.06347 (2017), 12.