Symbolic Reinforcement Learning for Safe RAN Control

Alexandros Nikou, Anusha Mujumdar, Marin Orlić and Aneta Vulgarakis Feljan
Ericsson Research, Sweden and India

Introduction

- Increased demand for self-organized and autonomous networks to address the growing complexity of modern cellular networks.
- Networks are required to ensure acceptable Quality of Service (QoS) to each user connected to the network.
- Reinforcement Learning is a promising solution for optimal decision and control of agents in an uncertain environment.
- Large-scale exploration performed by RL algorithms can lead to unsafe states.
- In this work, we demonstrate a novel approach for guaranteeing safety by applying model-checking techniques.

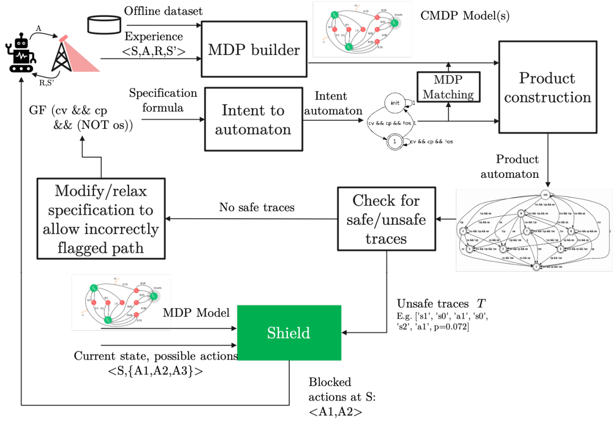
Contributions

- A general automatic framework taking user input in form of a LTL specification and deriving a policy that fulfils it.
- Blocking control actions that violate the Linear Temporal Logic (LTL) specification.
- Novel system dynamics abstraction to computationally efficient Markov Decision Process (MDP).
- User interface allowing the user to graphically access all the steps of the approach.

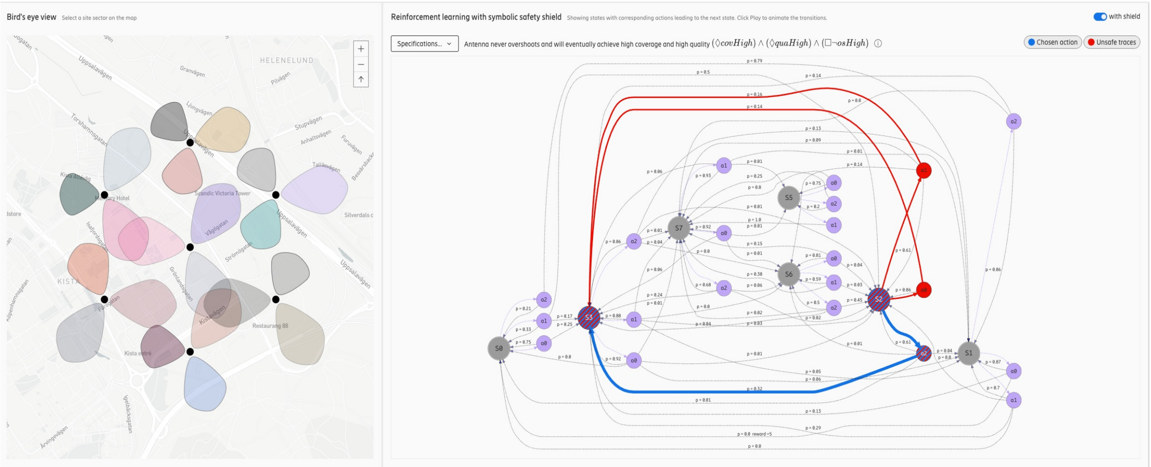
Applicability to other domains

- A general architecture that can be applied to any framework in which the dynamical system under consideration is abstracted into an MDP.
- Example other applications: robot planning with states of the MDP representing the state of the environment that the robot can move in. LTL tasks include both reachability and safety.

Demonstration



- The initial user intent, which can be written in LTL format is translated into an intent automaton.
- By gathering experience data tuples from the RL agent trained in simulation environment, we construct the system MDP.
- By computing the product between the MDP with the intent automaton, we have access to all system behaviors.
- By applying model checking and graph techniques, we are able to find the traces that violate the LTL task.
- If there exists some unsafe and safe traces the process moves to a shield strategy that blocks the actions that leads to unsafe traces.



Algorithm

Input: User specification Φ
Gather experience replay (s, a, r, s') from data;
Discretize states into N_b. State space size is $|\mathcal{S}|^{N_b}$;
Construct the MDP dynamics $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$;
Translate the LTL formula Φ to a BA $\mathcal{A}_\varphi$;
Compute the product $\mathcal{T} = MDP \otimes \mathcal{A}_\varphi$ and pass it to model checker;
Model checking returns traces that violate φ;
If no safe traces found **Modify/Relax** φ
Else Block unsafe actions by function *Shield*(MDP, $\mathcal{T}$).

Conclusions and future work

- We have demonstrated an architecture for network KPIs optimization guided by user-defined intent specifications given in LTL.
- Our solution consists of MDP system abstraction, automata construction, and model-checking techniques.
- Future efforts will be devoted towards applying the proposed framework in other telecom use cases as well as robot planning.

References

[1] M. Alshiekh. Safe reinforcement learning via shielding. Proceedings of the AAAI Conference on Artificial Intelligence , 2018.

[2] M. Bouton, et all. Point-Based Methods for Model Checking in Partially Observable Markov Decision Processes. Proceedings of the AAAI Conference on Artificial Intelligence (2020).

[3] S. Fan et all. Self-optimizationofcoverage and capacity based on a fuzzy neural network with cooperative reinforcement learning. EURASIP Journal on Wireless Communications and Networking 2014, 1 (2014), 57.

[4] V.Mnih, et al. 2015. Human-level control through deep reinforcement learning. nature 518, 7540 (2015), 529–533.

[5] F. Vannella et all. Off-policy Learning for Remote Electrical Tilt Optimization. arXiv preprint, arXiv:2005.10577 (2020).

Video link: https://www.ericsson.com/en/reports-and-papers/research-papers/safe-ran-control