ITU KALEIDOSCOPE ONLINE 2021

6-10 December 2021

Reinforcement learning for scheduling and MIMO beam selection using CAVIAR simulations

João Paulo Tavares Borges
Universidade Federal do Pará, Brazil

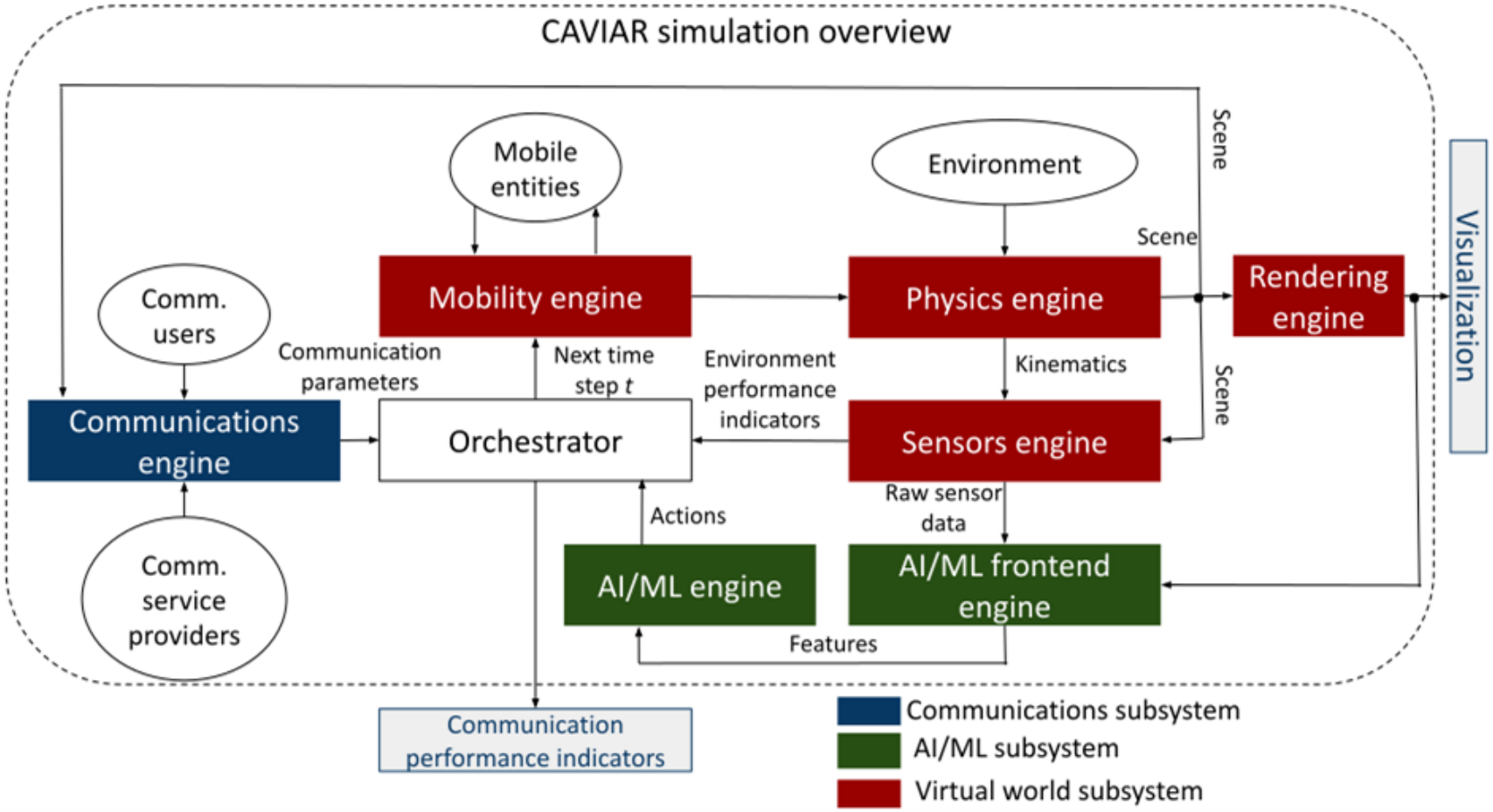
Session 6: Machine learning for next generation wireless network

Paper S6.3: Reinforcement learning for scheduling and MIMO beam selection using CAVIAR simulations

Introduction

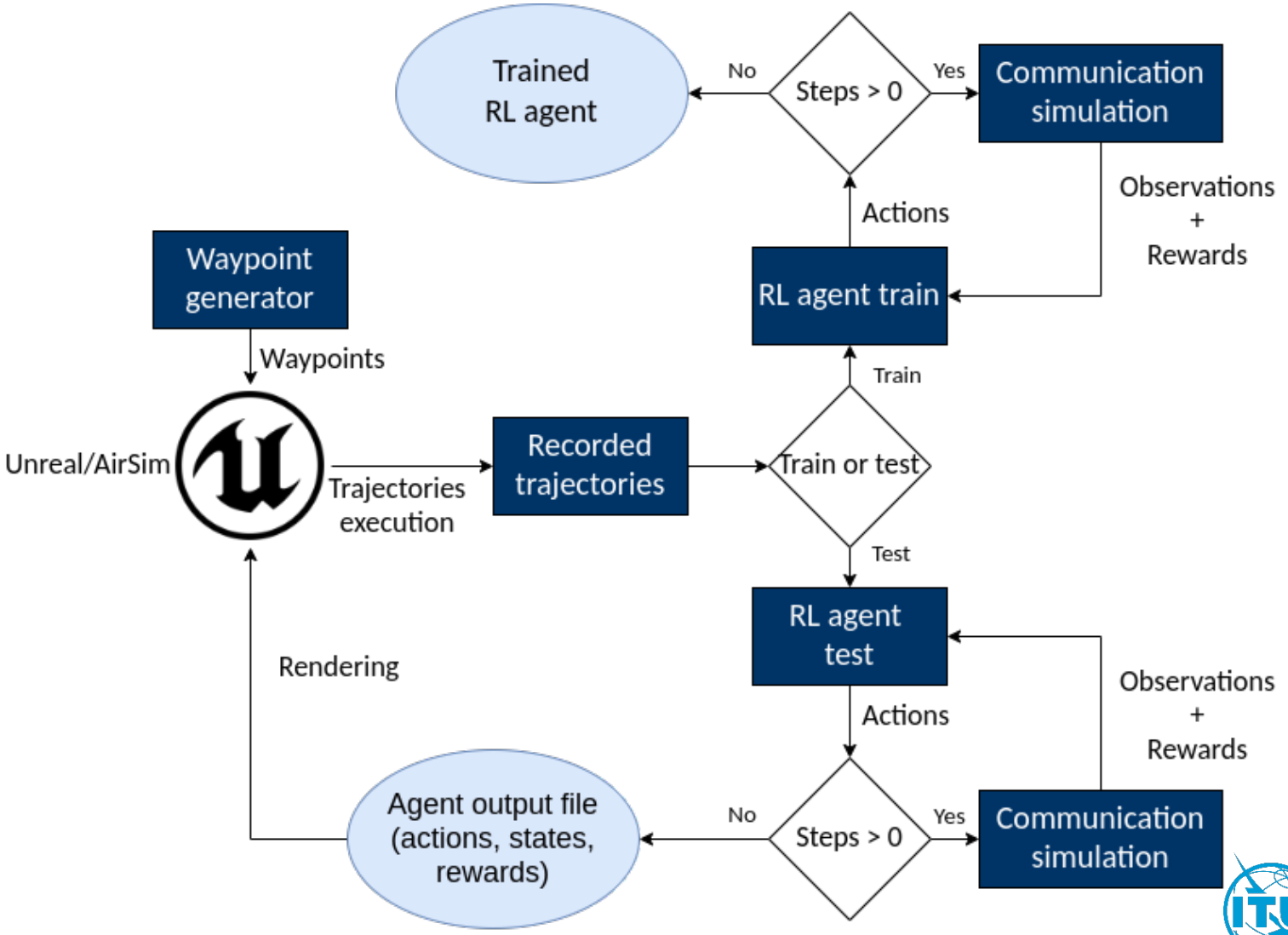
- ML and, more specifically RL, appears in **several research areas in 5G and 6G**
- However, **the lack of freely available datasets or environments to train and assess RL agents is a practical obstacle** that delays widespread adoption
- So, the key idea in this paper is to **use realistic representations of deployment sites together with physics, sensors, and communication network simulators**, to enable training RL agents
- This methodology is named **Communication Networks, Artificial Intelligence and Computer Vision with 3D Computer-Generated Imagery (CAVIAR)**

Overall description of a possible CAVIAR implementation



CAVIAR simulation for user scheduling and beam selection problems

- A set of mobile entities (UAVs, cars and pedestrians) **receive a set of waypoints** to follow
- While following the waypoints, **their trajectories are recorded** at a given sampling rate
- These **recordings are used, at each simulation step, to position the entities**, while the channels between them and the base station are calculated
- The communication parameters, together with UE movements, are **used to compose environments** (i.e. for user scheduling and beam selection)



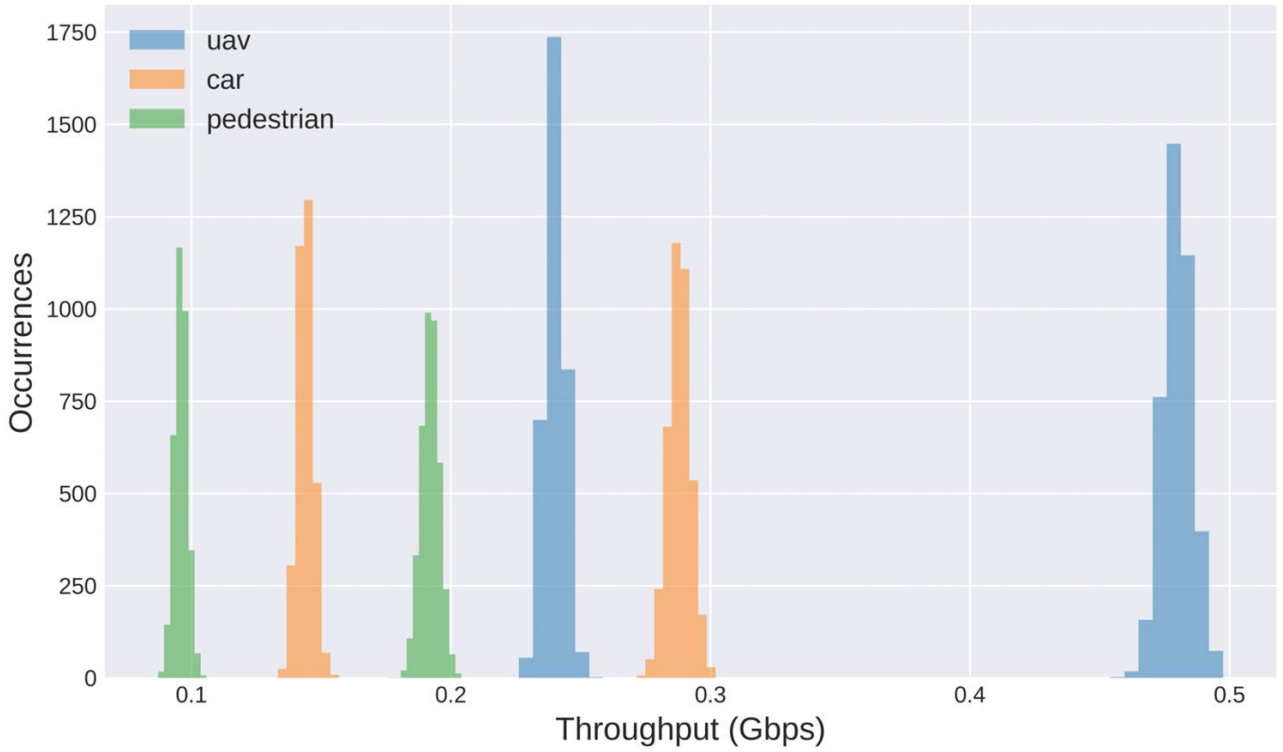
Environment communication parameters configuration

Fix beamforming codebook obtained from Discrete Fourier Transform (DFT) matrices for 8x8 Uniform Planar Arrays (UPA)

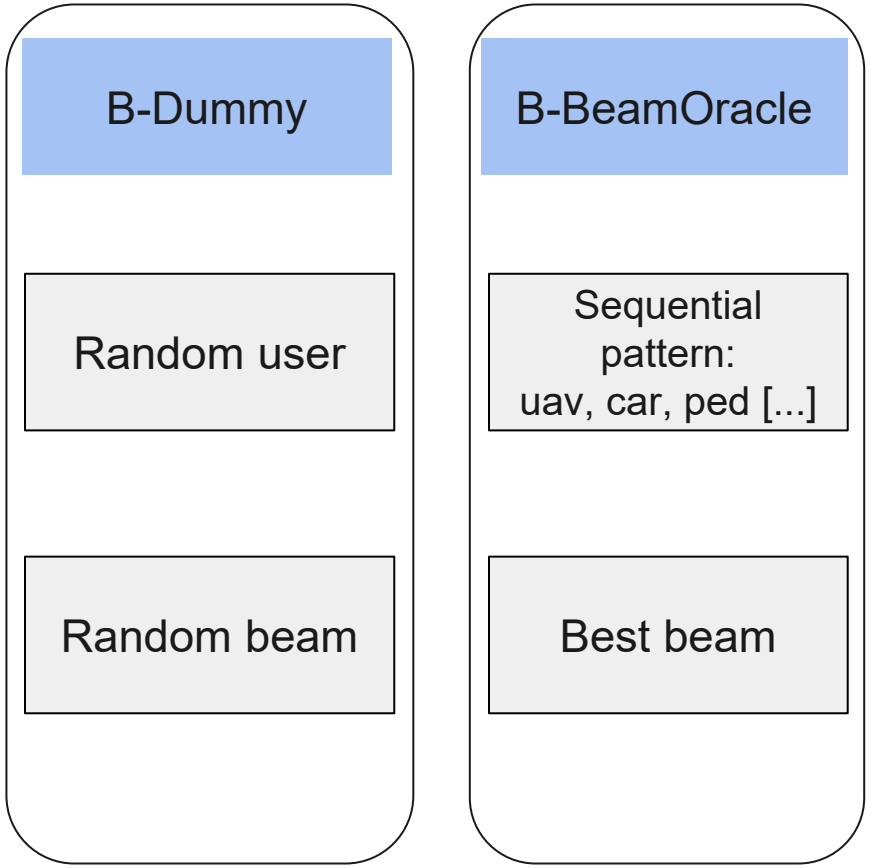
Three users with 1 Gbit buffer each, and traffic defined via a Poisson process

In total: 3 users and 64 beams, composing an action space of 192 actions

Network load	Total throughput	UAV (%)	Pedestrian (%)	Car (%)
Light	0.48 Gbps	50%	20%	30%
Heavy	0.96 Gbps	50%	20%	30%



Example RL agent and baseline strategies to test the environment



Baseline A2C agent (B-RL)

State:
X,Y,Z, dropped packets P_d , transmitted packets P_{tx} , buffered packets P_b , bitrate

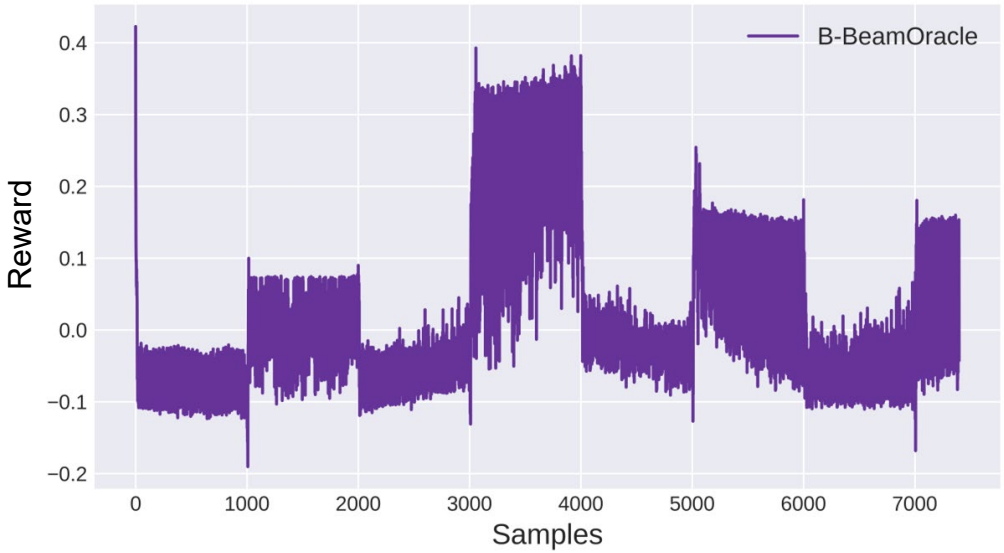
Action:
User (int), beam index (int)

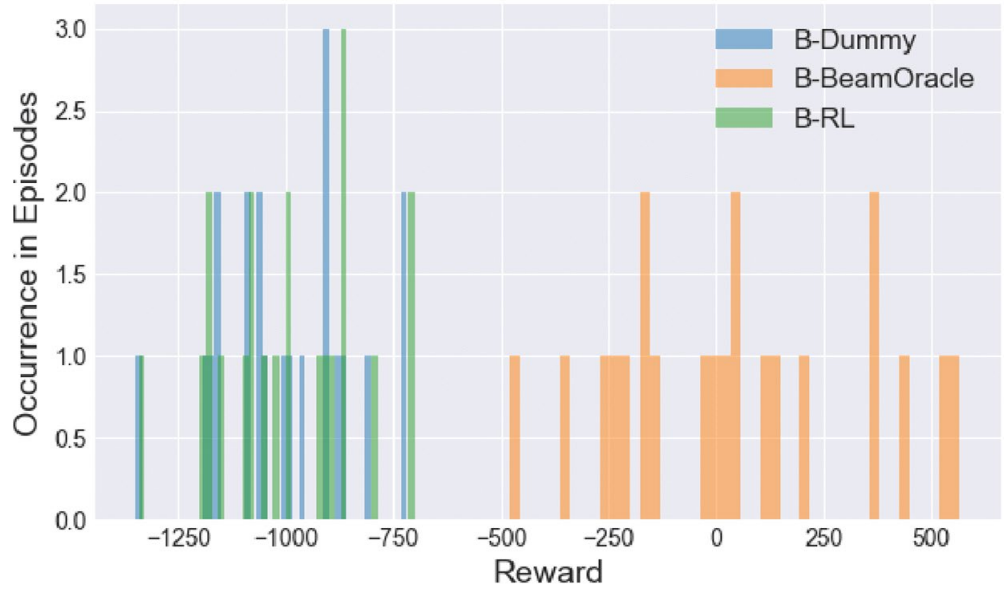
Reward:

$$r_e[t] = \frac{P_{\text{tx}}[t] - 2P_d[t]}{P_b[t]}$$

Results

- The CAVIAR environment was used to generate 70 episodes, from which 50 were used for training the RL agent, and 20 for testing
- The sequential scheduling proves to be sufficient to attend the demand in light traffic situations, however, for intense traffic moments, even using the best beam, without proper scheduling, the performance of the reward tends to be negative
- As expected, the B-BeamOracle presents the best performance, while the B-RL achieves performance close to the B-Dummy, which simply uses random actions
- We believe that the use of more episodes and better modeling of the agent can substantially improve its performance.





Conclusion and future works

- This paper presented a framework for research on RL applied to scheduling and MIMO beam selection
- Using the framework, we provided statistics of an experiment in which an RL agent faces the problems of user scheduling and beam selection
  - the experiment allowed us to validate the designed environment for RL training and testing
- Future development will focus on
  - in-loop simulations, where there is no need to generate trajectories in advance
  - adding integration with more simulators to increase realism, such as a full-fledged network traffic and a wave propagation with ray tracing software

ITU KALEIDOSCOPE
ONLINE 2021

Thank you!