The CommsRLTimeFreqResourceAllocation-v0 environment

Allocate radio resources to UEs.

On each episode of this environment, the agent must allocate N_f downlink frequency resources to User Equipments (UEs). This takes place in a free-space scenario with K UEs, where each UE has specific traffic requirements (some require high guaranteed bit rates, others low packet delivery delays, etc.). This recreates a well-known case of OFDM resource allocation, where a MAC scheduler allocates frequency resources to UEs under different radio conditions.

MDP dynamics

At the beginning of an episode, K UEs are scattered randomly throughout an empty Euclidean space containing a BTS at coordinates (0, 0). The BTS transmits with EIRP=13 dBm and the space is of size 1 km^2 centered around the BTS. The carrier frequency is $f_{carrier} = 2655$ MHz and the system bandwidth $BW = 5$ MHz. The transmit power is distributed equally across all PRBs. Free space propagation is assumed and the UEs move at random speeds in random rectilinear trajectories throughout the environment (bouncing off the edges at specular angles). The UE speeds are normally distributed as described in Table 2 of [1] for *Overall* pedestrians in Location 2.

Each episode begins at time step $t = 0$ with $p = 0$ and TTI=0, where p denotes the current PRB being allocated and TTI is the Transmission Time Interval. One TTI is assumed to last 1 ms exactly. The environment is then time-stepped and the TTI counter is increased by 1 when $p = t \ mod \ N_f = 0$. The environment is run indefinitely (i.e. for a very large number of time steps).

When the environment starts, each UE gets assigned a random QoS Identifier (QI) class from a total of 4 QIs. This assignment is uniform (There are exactly K/4 UEs of the same QI and all QIs are assigned).

On the first time step of each TTI (i.e. when $t \ mod \ N_f = 0$), the environment generates (or not) new traffic packets for each UE according to their QoS Identifier class (see Table 1 below). These packets are then added to the UE's traffic buffer. A packet size in a UE's buffer decreases each time step according to the UE's spectral efficiency and to the number of radio resources allocated by the agent to the UE. The maximum number of packets that each buffer can store (i.e. its buffer size) is defined as $L = 100$.

Observation space

The state vector is a concatenation of vectors providing the following information at each time step:

- Channel Quality Indicator (CQI) of each and all UEs. $q_k \in [0,15] \ \forall k \in [1, K]$
- Sizes (in bits) of all packets in each UE's buffer. $S = (s_{k,l}) \in \mathbb{R}^{KxL}$, where $s_{k,l}$ is the size of the l^{th} packet of the k^{th} UE. S is flattened in row-major order in the state vector.
- Ages (in TTIs) of all packets in each UE's buffer. $E = (e_{k,l}) \in \mathbb{R}^{KxL}$, where $e_{k,l}$ is the age of the l^{th} packet of the k^{th} UE. E is flattened in row-major order in the state vector.
- QoS Identifier (QI) classes of each and all UEs as a one-hot vector. $c_k \in [0,1,2,3] \ \forall k \in [1, K]$. The QI classes are given in Table 1.

Table 1 : Traffic characteristics of each QoS Identifier class

QI classes	Resource Type	GBR [kbps]	Packet Delay Budget [ms]
3 or [0,0,0,1]	GBR (*Conversational V*oice)	29.2	100
2 or [0,0,1,0]	GBR (Conversational Video)	1250	150

1 or [0,1,0,0]	Delay Critical GBR	10	30
0 or [1,0,0,0]	Non-GBR (web browsing)	N.A.	300

- Current PRB being allocated $p \in [0, \dots, N_f - 1]$, which can be calculated from the current timestep as $p = t \ mod \ N_f$.

Action space

On each time step, the agent may take any of K possible actions, thus assigning the current frequency resource to the k[th] UE. If an action is chosen that allocates the current PRB to the k[th] UE, the size of the oldest packet in the k[th] UE's buffer is reduced by a number of bits equal to the number of transmitted bits. The number of bits transmitted in one PRB depends on the UE's channel quality.

Reward

The agent receives a reward of 0 on all time steps except on those leading to a state wherein $p = 0$. These are called TTI transition time steps.

The reward received on the TTI transition time steps is the negative sum of non-GBR buffer sizes (to encourage the agent to empty the non-GBR queues as fast as possible), plus the negative sum of delay traffic (to encourage the agent to respect the Packet Delay Budgets). Note that this reward is calculated before new traffic is added to the UEs' buffers:

$$r_t = r_t^{(GBR)} + r_t^{(nonGBR)}$$

Where

$$r_t^{(GBR)} = -\sum_{k=1}^{K} \sum_{\substack{l=1 \\ e_{k,l} > PDB_k \\ c_k \in [1,2,3]}}^{L} s_{k,l}$$

$$r_t^{(nonGBR)} = -\sum_{k=1}^{K} \sum_{\substack{l=1 \\ e_{k,l} > PDB_k \\ c_k = 0}}^{L} s_{k,l} - \sum_{k=1}^{K} \sum_{\substack{l=1 \\ c_k = 0}}^{L} s_{k,l}$$

Default parameters

Parameter	Value
L	100
N_f	10
K	128
$f_{carrier}$	2655 MHz
BW	5 MHz

References

[1] S. Chandra and A. K. Bharti, "Speed Distribution Curves for Pedestrians During Walking and Crossing," *Procedia - Soc. Behav. Sci.*, vol. 104, pp. 660–667, 2013.