CS 285–Fall 2023 — Project Proposal

1 Online Optimization comparison with RL

2 Reinforcement Learning Formulation of the Aircraft Service Problem

We define the aircraft service problem as a single-agent Markov decision process. The environment consists of a fleet of N aircrafts $\mathbb{A}_i$ and a set of $M = 2$ vertiports $\mathbb{V}_i$. These aircrafts serve passengers moving between the two vertiports.

2.1 Observation

This observation contains information about all aircraft. For each aircraft i at time t, the state $s_t^{(i)}$ includes $[\text{position}_t^{(i)}, \text{SOC}_t^{(i)}, \text{condition}_t^{(i)}, \text{high-level action}_t^{(i)}, \text{low-level action}_t^{(i)}, \text{action duration}_t^{(i)}]$.

2.2 Action

We adopt a hierarchical action space.

- **High-level Actions:**
  1. **Keep Charging:** The aircraft stays at its current vertiport and continues charging.
  2. **Go for a Flight:** The aircraft flies to the other vertiport.
- **Low-level Actions:** Define specifics for how long an aircraft should charge, or the speed/intensity of the flight to the other vertiport. For instance, for the "Keep Charging" action, the continuous low-level action could be the duration T_C for which the aircraft should keep charging.

Therefore, our action space is defined as $\{k, X_k\}$ where $k \in \{1, 2\}$ denotes one of the two possible discrete actions and $X_k \in [0, T_{\text{threshold}}]$ denotes the continuous part, which will be executed if needed.

2.3 Reward

The primary objective of our aircraft service strategy is to ensure minimal passenger waiting times and energy consumption, while also equalizing aircraft condition. Thus, our reward function is defined as:

$$r = -(\alpha \times \text{Total Waiting Time} + \beta \times \text{Energy Consumed} + \gamma \times \text{Aircraft Condition Disparity})$$

Where:

- Total Waiting Time is the cumulative waiting time of all passengers.
- Energy Consumed is the total energy used by the aircraft during the flight.
- Aircraft Condition Disparity is a measure of how unevenly the aircraft conditions are being utilized.
- α, β, and γ are weights to balance the significance of each factor in the reward function.

In practice, there may be additional considerations or modifications needed based on detailed specifications of the problem or based on preliminary results from training the RL agent.