

Hardware and Implementation Details

The planner in our TeamWeaver system is based on the Llama-3-8B-Instruct model, operating with a temperature of 0 and a token limit of 2048, and is integrated with an MIQP solver.

All experiments were executed on two workstations, one equipped with an RTX A6000 and the other with dual RTX 4090D GPUs, each supported by a 32-core CPU under a fixed simulation budget. To ensure fair comparisons, all evaluations use the same simulation settings and hyperparameters.