INDIAN INSTITUTE OF INFORMATION TECHNOLOGY DHARWAD

DeepCoMP: Self-Learning Dynamic Multi-Cell Selection for Coordinated Multipoint (CoMP)

CELLULAR MOBILE COMMUNICATION MINI PROJECT REPORT

Submitted by :
Abhishek Singh Kushwaha (19BEC001)
Sriram Shivganesh (19BEC043)

Under the guidance of :
Dr. Mukesh Kumar Mishra
Assistant Professor, Department of Electronics & Communication Engineering, IIIT Dharwad

PREFACE

The work presented in this project report results from our mini-project work at the Indian Institute of Information Technology Dharwad. The research and experimentations were conducted between January 2022 to April 2022 under the guidance of Dr. Mukesh Kumar Mishra, Assistant Professor, Faculty in Electronics & Communication Engineering@IIIT Dharwad.

In this project, we have implemented a tool called DeepCoMP, which is used to allocate the multiple coordinated point connections in the 5G and beyond a generation. It uses reinforcement learning to allocate resources and maximize the quality of experience for users.

Abhishek Singh Kushwaha
Sriram Shivganesh

Dharwad, May 2022

ACKNOWLEDGEMENT

Our sincerest gratitude goes out to our advisor, Dr. Mukesh Kumar Mishra, for allowing us to carry out this mini-project and mentoring and advising us throughout this semester-long mini-project.

We would like to thank the whole department of Electronics & Communication Engineering at the Indian Institute of Information Technology Dharwad for providing me with an opportunity to work with them and contribute to their research and development.

We would also like to thank our families and friends for encouraging us and giving us moral and emotional support during the transition from online mode to offline mode.

Abhishek Singh Kushwaha
Sriram Shivganesh

Dharwad, May 2022

TABLE OF CONTENTS

What is Reinforcement Learning

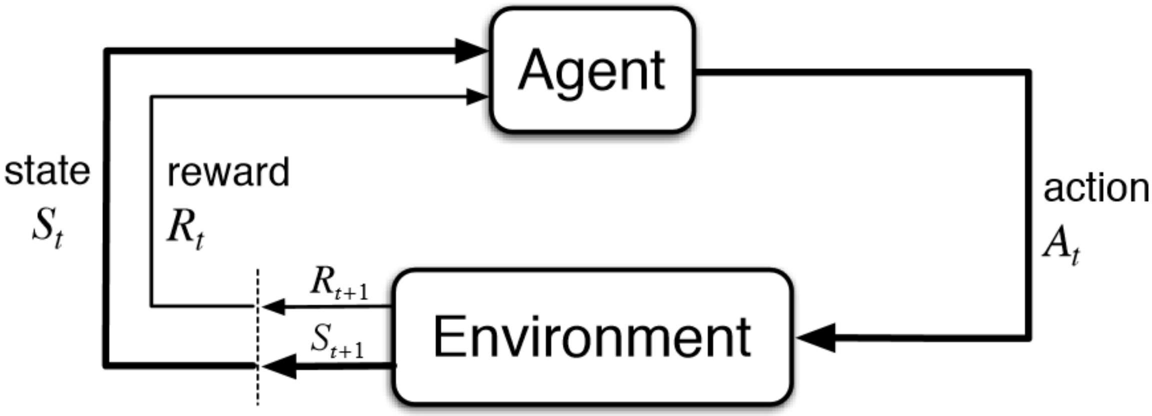
Reinforcement Learning is a feedback-based Machine learning technique in which an agent learns to behave in an environment by performing the actions and seeing the results of actions.

For each good action, the agent gets positive feedback or reward, and for each bad action, the agent gets negative feedback or a penalty. The agent interacts with the environment and explores it by itself. The primary goal of an agent in reinforcement learning is to improve performance by getting the maximum positive rewards. We do not need to pre-program the agent, as it learns from its own experience without any human intervention.

RL solves a specific type of problem where decision-making is sequential, and the goal is the long-term solution.

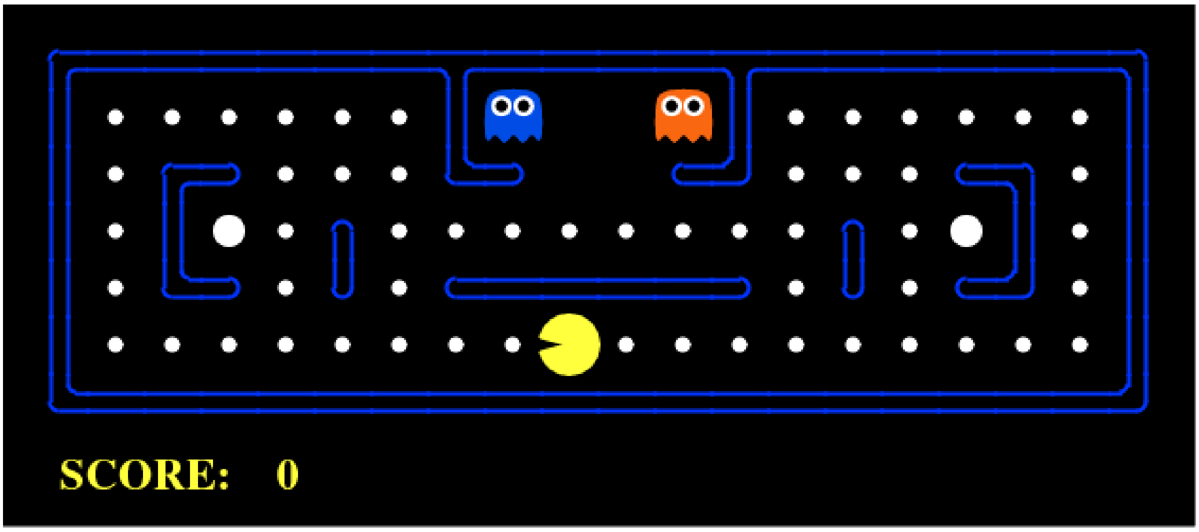
Current use cases include the following:

A. resource management
B. gaming
C. robotics



Example of Reinforcement Learning :

Using Reinforcement Learning, the Pac-Man game was fully automated and there was no need for players to sit and play.



CoMP- Coordinated Multipoint

Coordinated MultiPoint (CoMP) is based on transmission and reception at multiple separated sites with dynamic coordination among them.

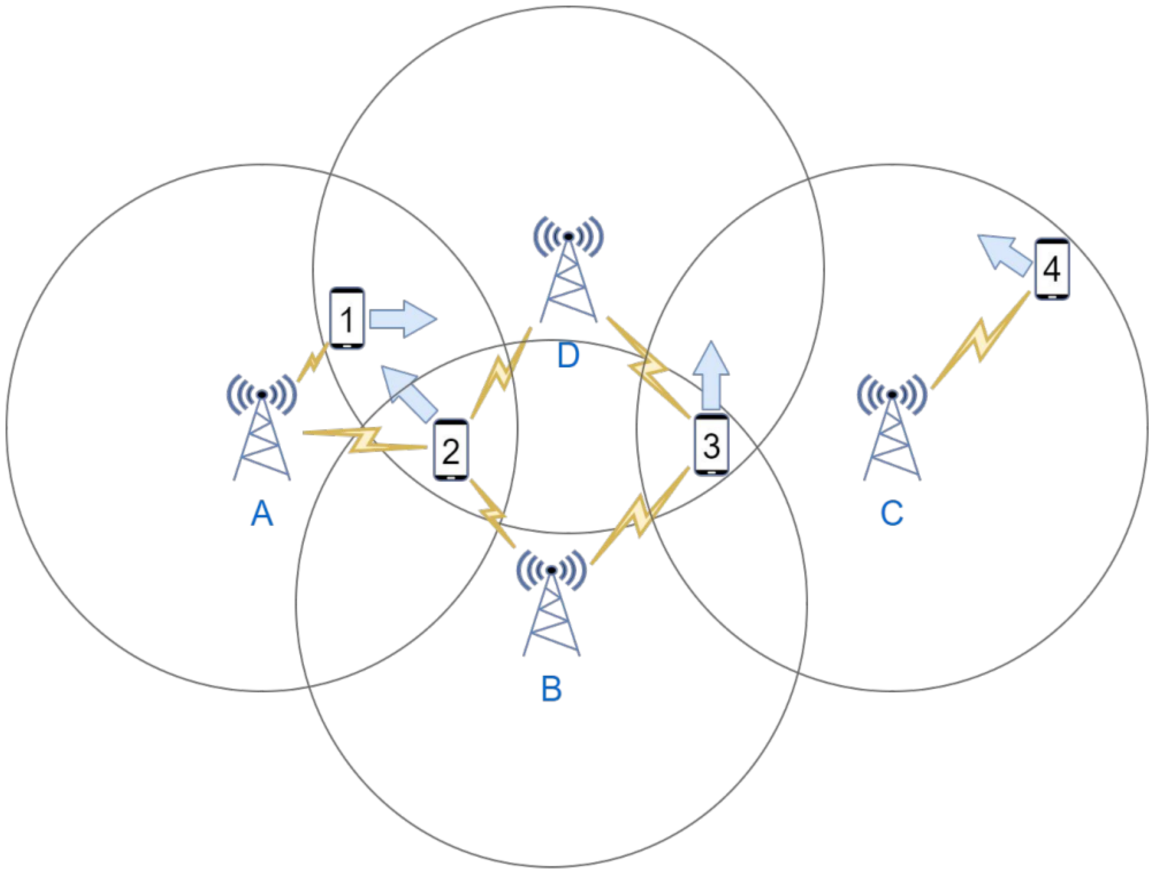
CoMP started to be used more aggressively in LTE Advanced, as a way of improving service at the cell edge.

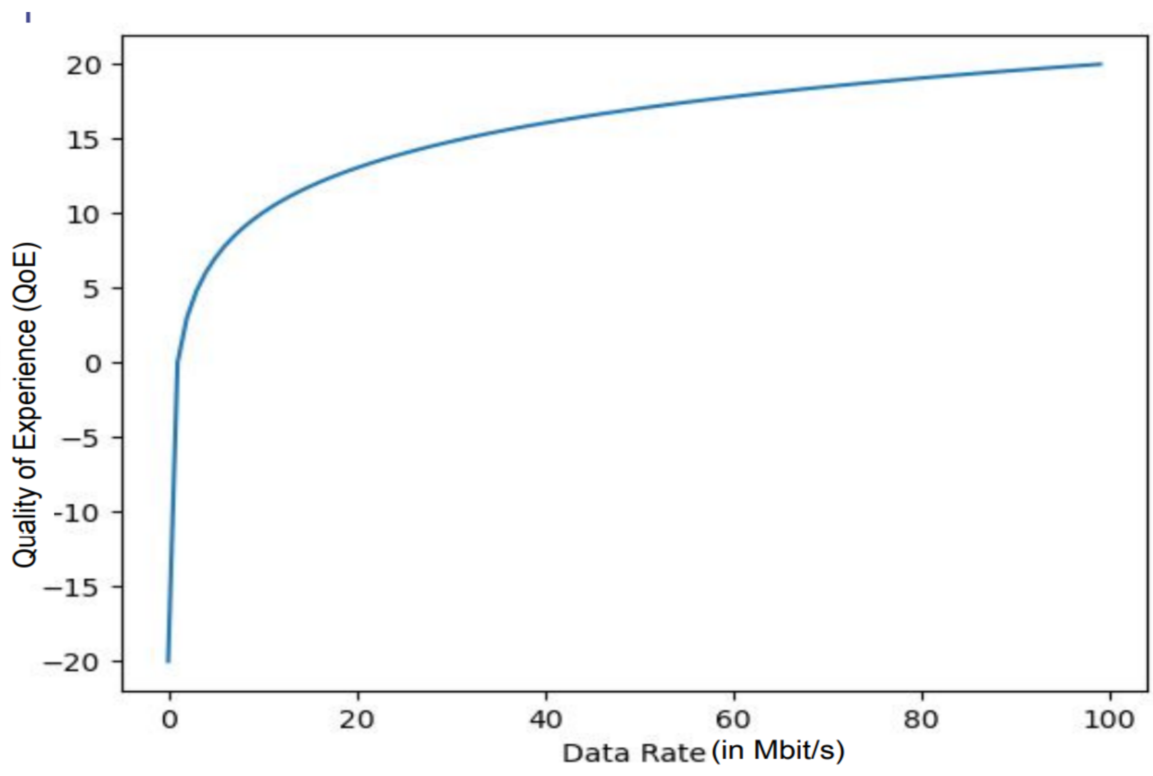
Makes better utilization of the network by providing connections from several base stations at once.

Problem Statement

DeepCoMP is our approach for multi-cell selection in 5g and beyond using reinforcement learning.

So, let's consider a use case where we consider a wireless mobile scenario where we have dense and partially overlapping cells in the image indicated by the antennas and the circles indicate the rough range of these cells, and then we have users moving around these cells represented by the smartphones. These users want services like AR, VR, cloud gaming, and video streaming all things that require high and reliable data rates, and one option or one way to achieve such high reliable data rates is using coordinated multipoint. Here we focus on coordinated scheduling and joint transmission where users can connect to not just a single cell but to multiple cells at once and then receive data from these multiple cells simultaneously such that their effectively received data rate is significantly increased. As these users move around and connect to the different cells, they, of course, compete for limited radio resources for example in the form of radio resource blocks which are limited in frequency and time domain. Our goal is to optimize dynamic multi-cell selection, for each user over time, we want to decide how many cells each user should connect to and which cells it should connect to. The goal here is to maximize the total quality of experience of all of our users. We model this quality of experience based on some previous studies as a logarithmic function of the data rate for each user. If the data rate is zero then the quality of experience is really bad. This utility function encourages fairness because if we want to maximize the total quality of experience for all of our users, we can do this most effectively by improving the data rate of users that currently have a low data rate whereas improving the data rate of users that already have a high rate is not so helpful. So, our goal is to maximize this total quality of experience and we want to do this with deep reinforcement learning.


1
4
D
2
3
A
C
B


20
15
10
5
0
−5
-10
−15
−20
Quality of Experience (QoE)
0
20
40
60
80
100
Data Rate (in Mbit/s)

Approaches to Solve This Problem

We have three different approaches:

A. DeepCoMP
B. D3-CoMP
C. DD-CoMP

- ## DeepCoMP

DeepCoMP is the centralized approach. It requires a global view and global control of all of our users in the area. This is a bit challenging from a technical perspective but it also means that we get a large observation of large action space. Has a high complexity and training effort which is a bit of the downside of this approach but on the upside, it can learn really powerful and fine-grained coordination and cell selection policies where the different users cooperate with each other because it does observe and can simultaneously control all of the users together.

The observations are based on the current connections of each user to the different cells. The signal strength between the different cells and users. Based on the user's current quality of experience. So given such an observation, the DeepCoMP agent makes an action. The action defines the cell selection for all of our different users. First is keeping the connections as it is, the agent can also decide whether a user should connect to a new cell or should disconnect from an existing cell or from a connected cell. The action space here is designed in a way that each user can make at most one connection or disconnection per time step simply to limit protocol overhead. With these observations and actions, the agent after each action receives a reward which we define to simply be the sum of all users' quality of experience because that's exactly what we want to optimize.

- ## D3-CoMP

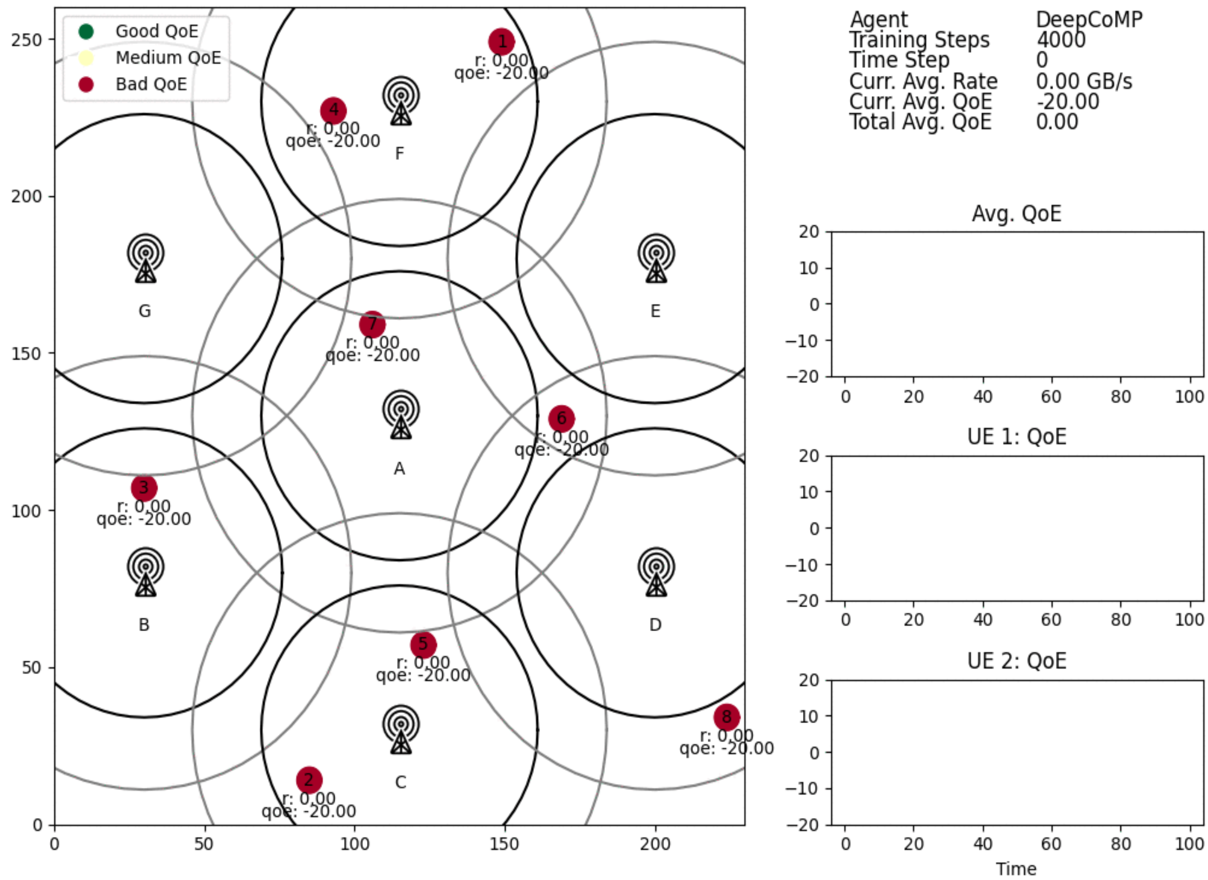
D3-CoMP is fully distributed where we have these independent DRL (deep reinforcement learning) agents each with its own separate policy and neural network. They all have their own policy that's being trained independently. They can potentially learn heterogeneous cell selection policies.

- ## DD-CoMP

DD-CoMP is a multi-agent approach. It also relies on local observations and actions but here these agents at least during training use a shared policy and a shared neural network and all the agents combine their experience during training such that we have more data than we can leverage during training from all the different users and it also means that we can often learn better and slightly more robust policies than with the D3-CoMP.

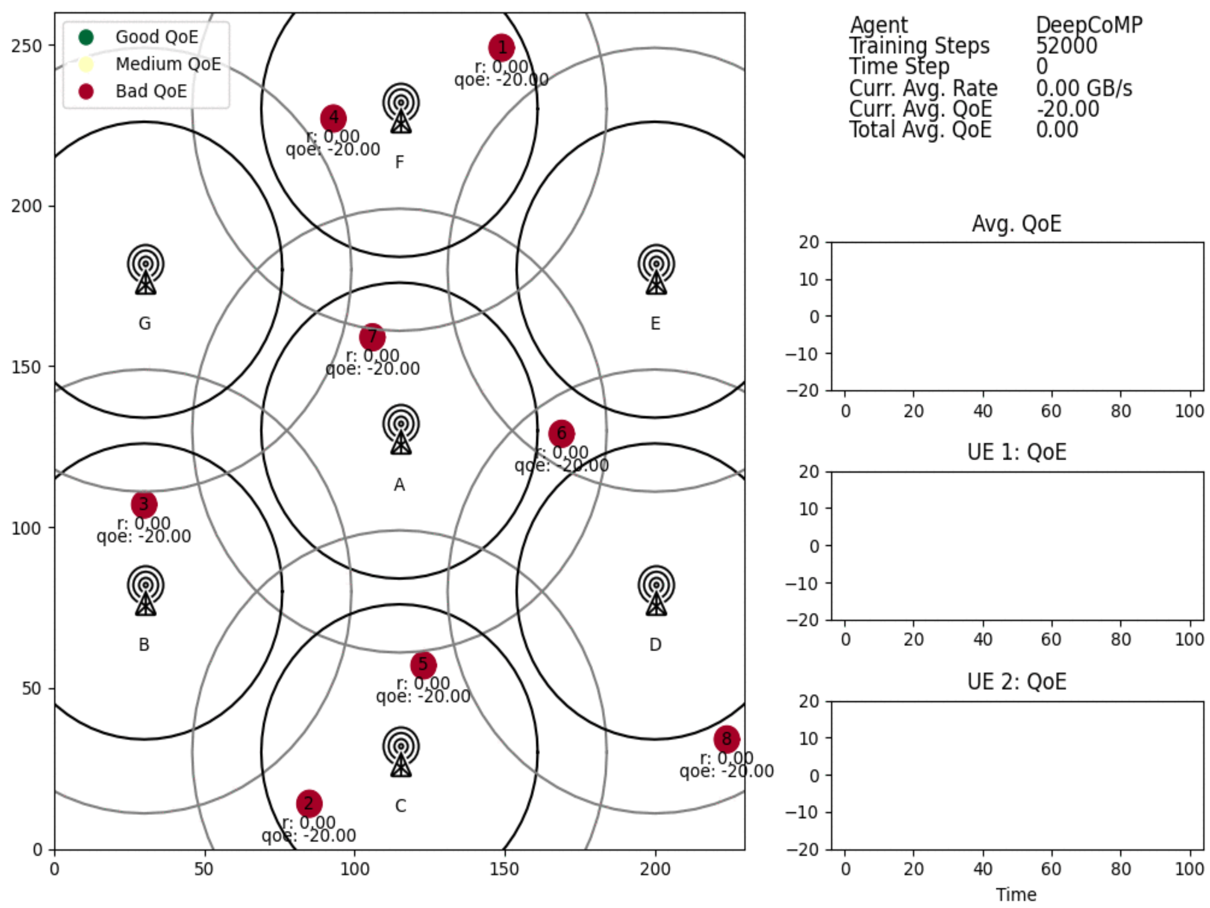
Results

- **DeepCoMP**



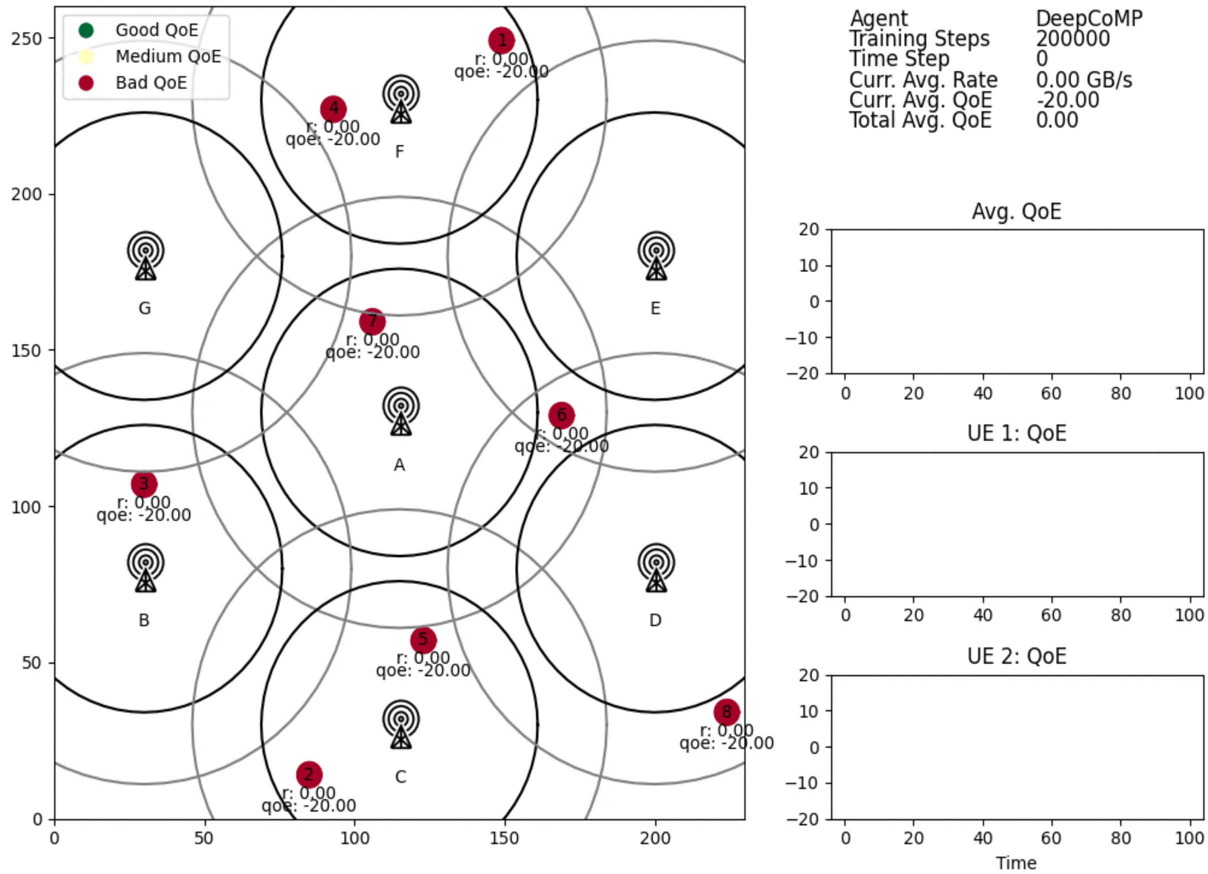
Training Steps: 500

Quality of Experience: Bad



Training Steps: 50,000

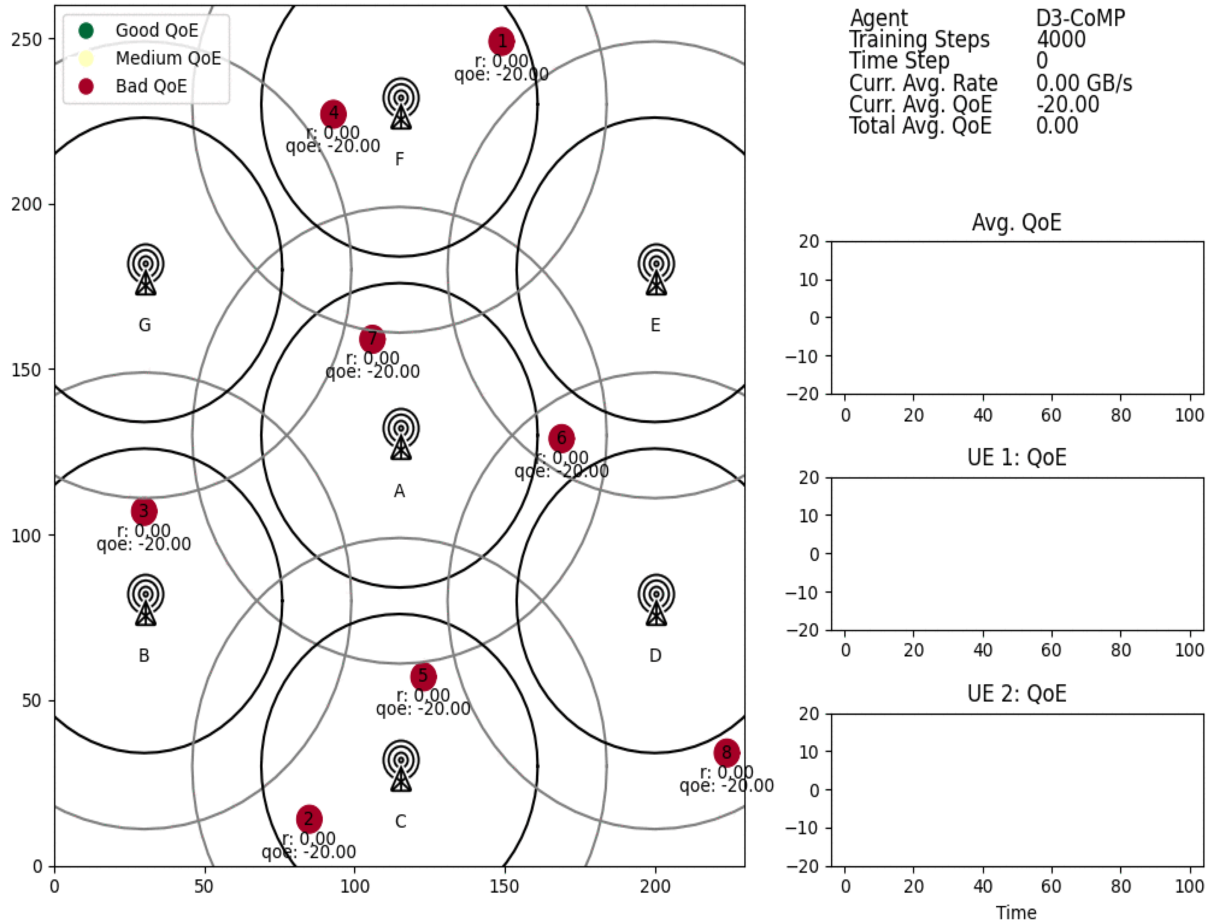
Quality of Experience: Decent



Training Steps: 2,00,000

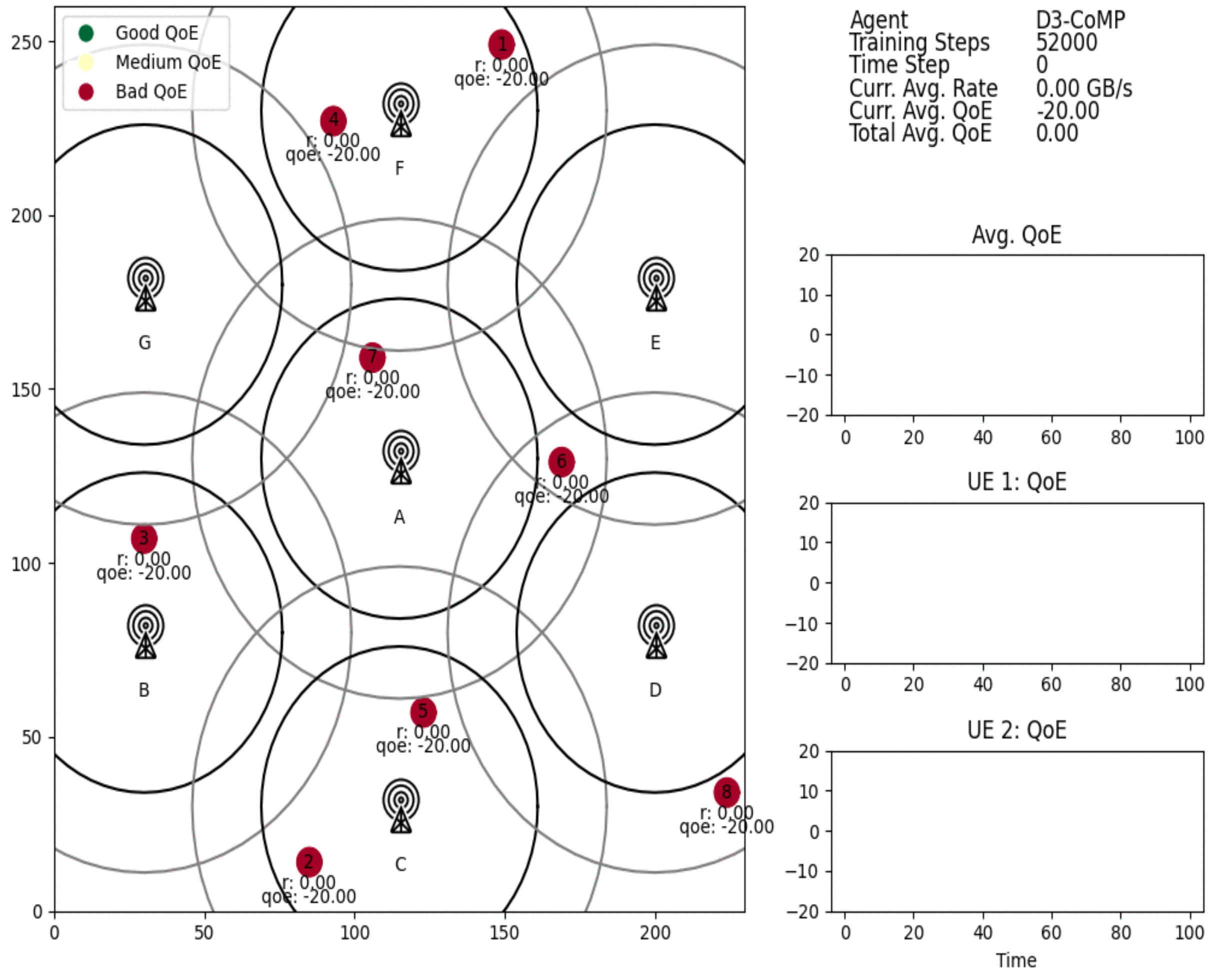
Quality of Experience: Good

• D3-CoMP



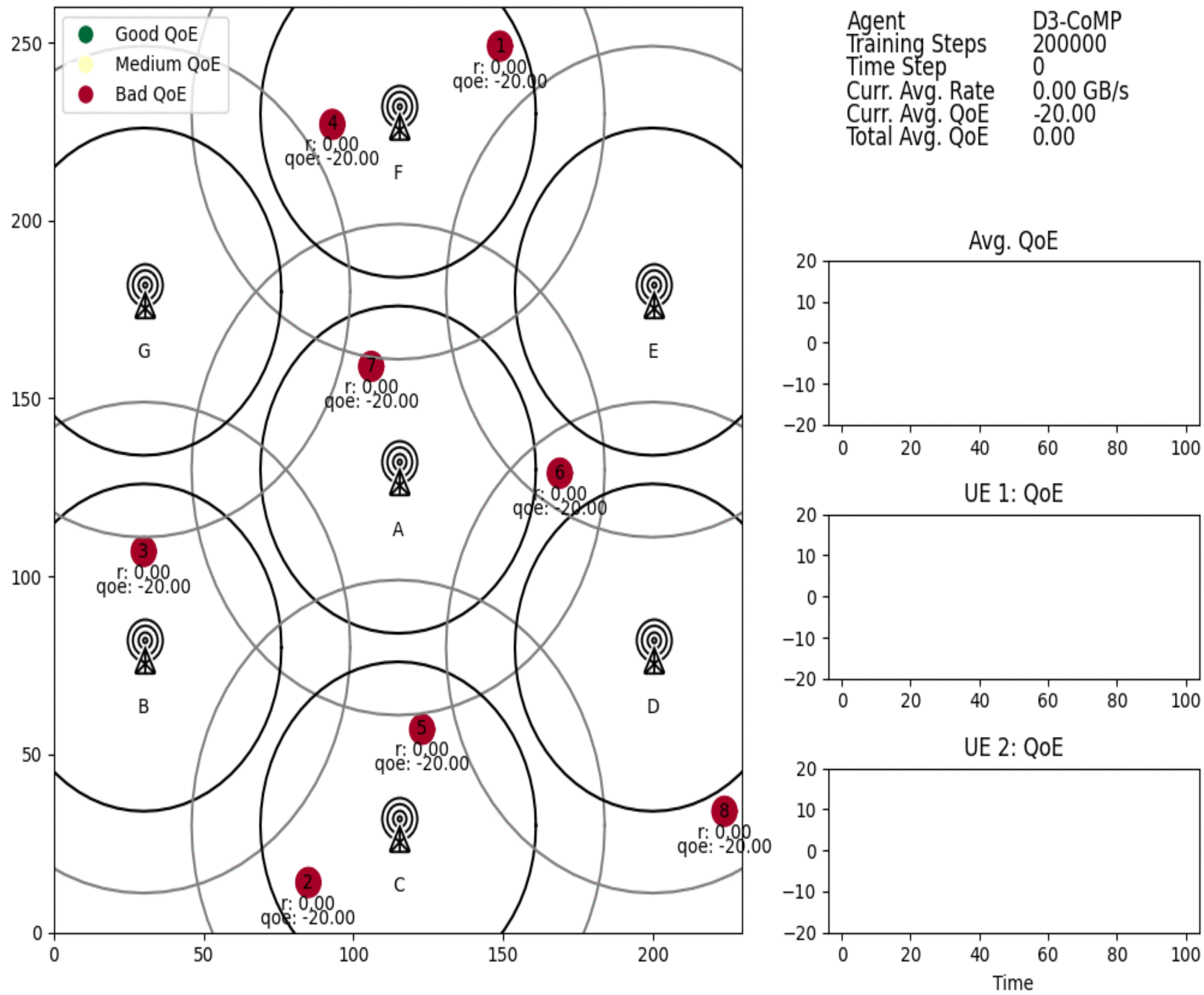
Training Steps: 500

Quality of Experience: Bad



Training Steps: 50,000

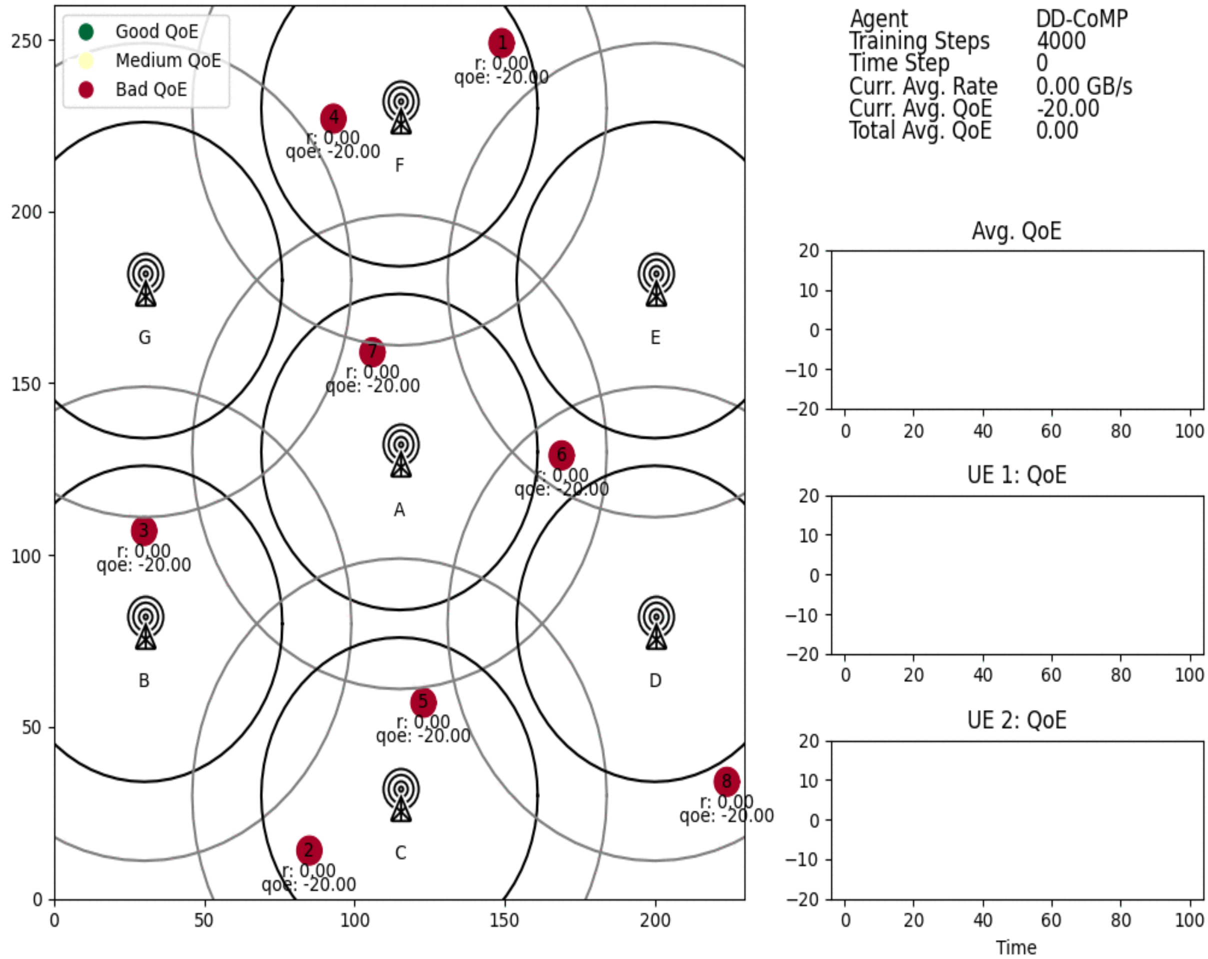
Quality of Experience: Decent



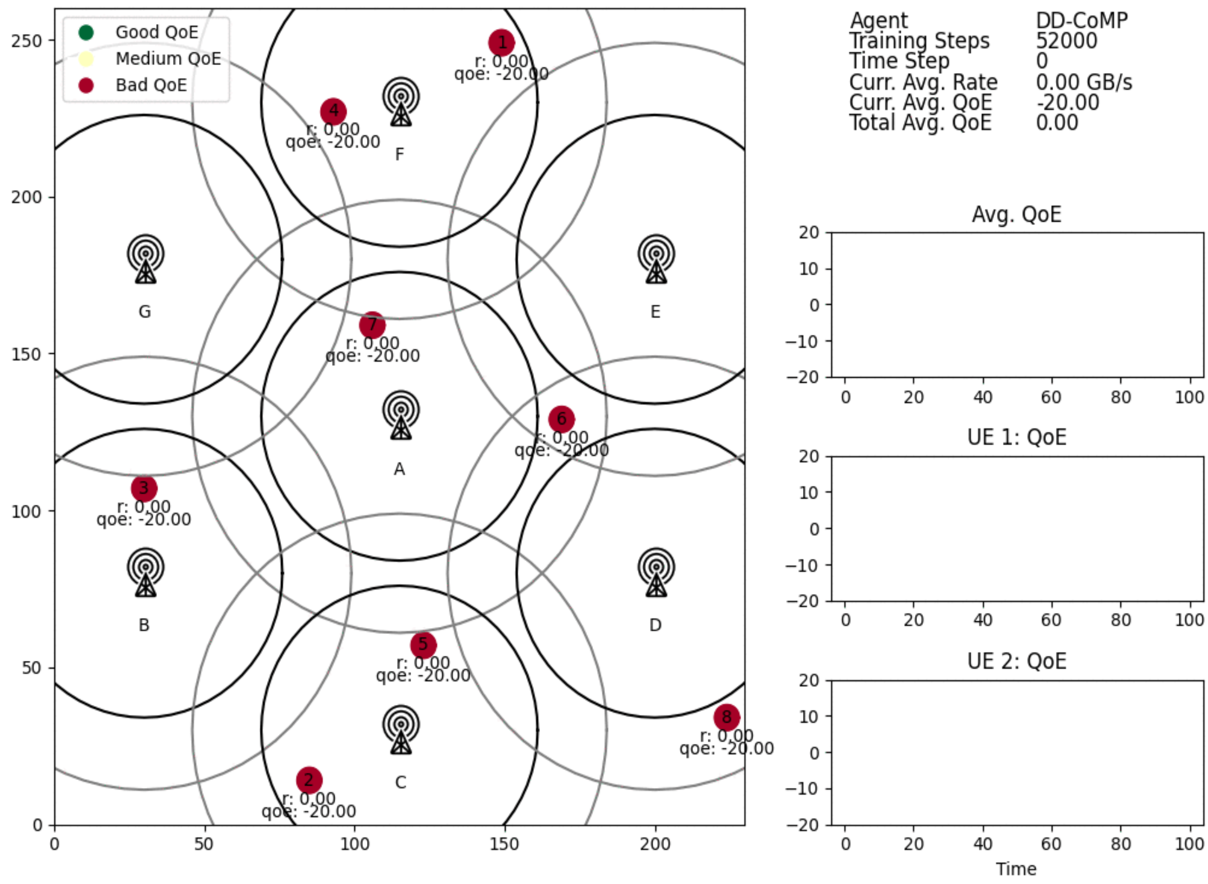
Training Steps: 2,00,000

Quality of Experience: Good

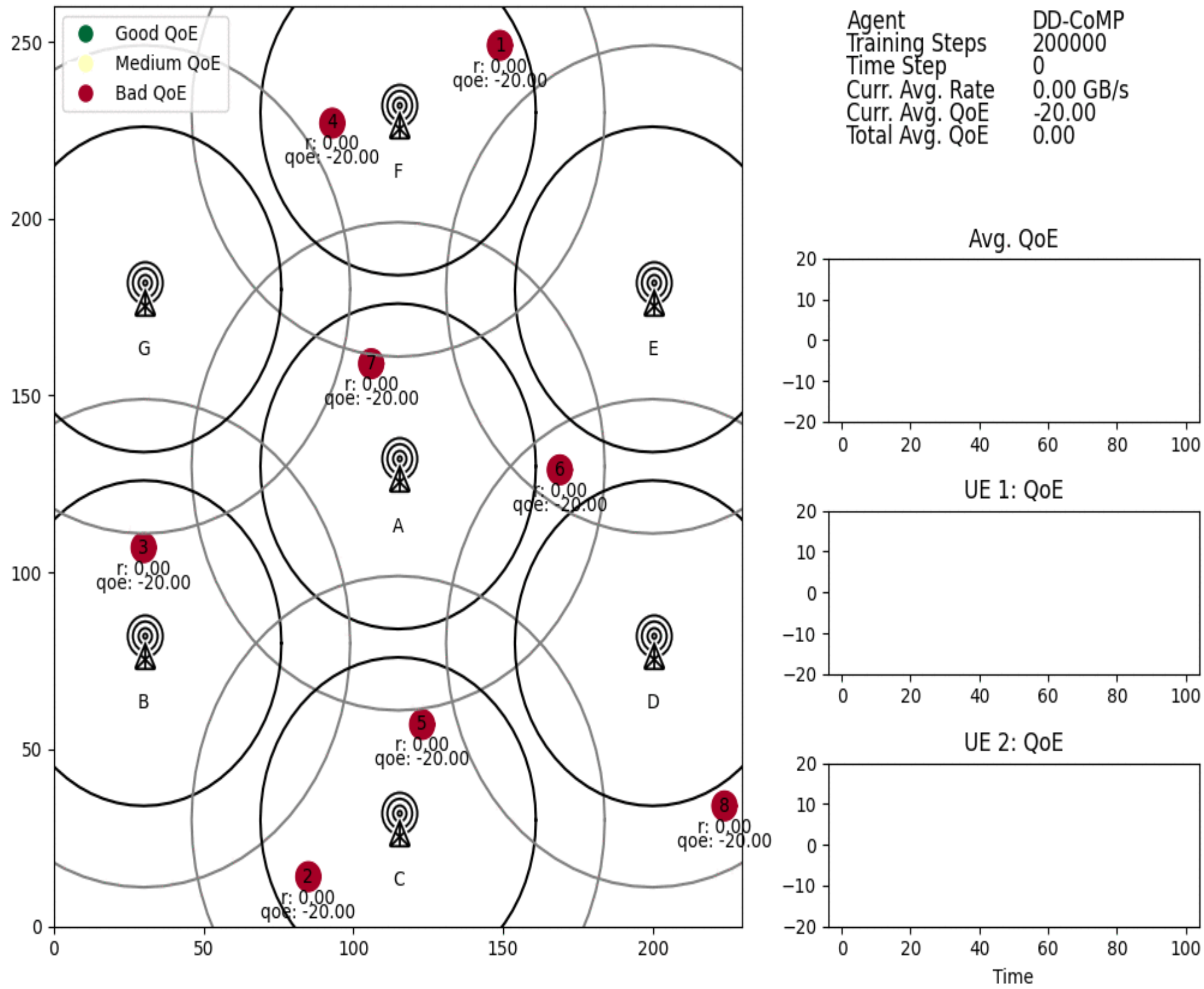
- **DD-CoMP**



Training Steps: 500
Quality of Experience: Bad



Training Steps: 50,000
Quality of Experience: Decent



Training Steps: 2,00,000

Quality of Experience: Good

References

- Stefan Schneider, DeepCoMP: Self-Learning Dynamic Multi-Cell Selection for Coordinated Multipoint (CoMP); Link
- Stefan Schneider, Prof. Dr. Holger Karl, Network and service coordination: conventional and machine learning approaches; Link
- Stefan Schneider, Ramin Khalili, Adnan Manzoor, Haydar Qarawlus, Rafael Schellenberg, Holger Karl, Self-Learning Multi-Objective Service Coordination Using Deep Reinforcement Learning; Link
- Stefan Schneider, Adnan Manzoor, Haydar Qarawlus, Rafael Schellenberg, Holger Karl, Ramin Khalili, Artur Hecker, Self-Driving Network and Service Coordination Using Deep Reinforcement Learning; Link