Updated Thesis

Alkis Sygkounas

December 2021

Thesis Update

1 Introduction

In this update, we introduced some new elements that they improved the current model. Specifically, there 's been a change on the structure of the main neural network, an introduction of a new algorithm for pathing with the usage of a new neural network and a new neural network called switcher that chooses in each time step between of those nets. In general, there 's been an increase in total coverage while almost reaching the 100% and also reduction in the oscillations for each different testing episode.

2 New Components

Firstly, the main neural net has changed to a deeper model. Specifically, the total number of hidden layers were changed from 3 to 7 and the total number of neurons in each hidden layer has been reduced to 256 from 512. The input and output size remained the same. So the total number of parameters (weights) reduced from 535.040 to 333.829 (37% decrease total) but total coverage gain was around 10% (median of 60% went to 70%). Still we see scenarios with extreme low coverage (40%).

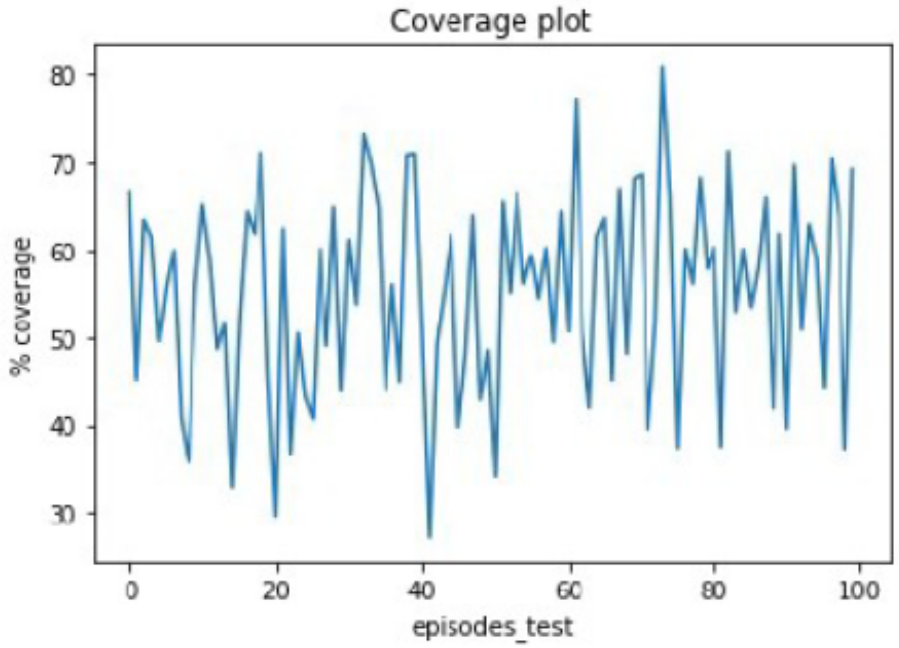

Figure 1: old model

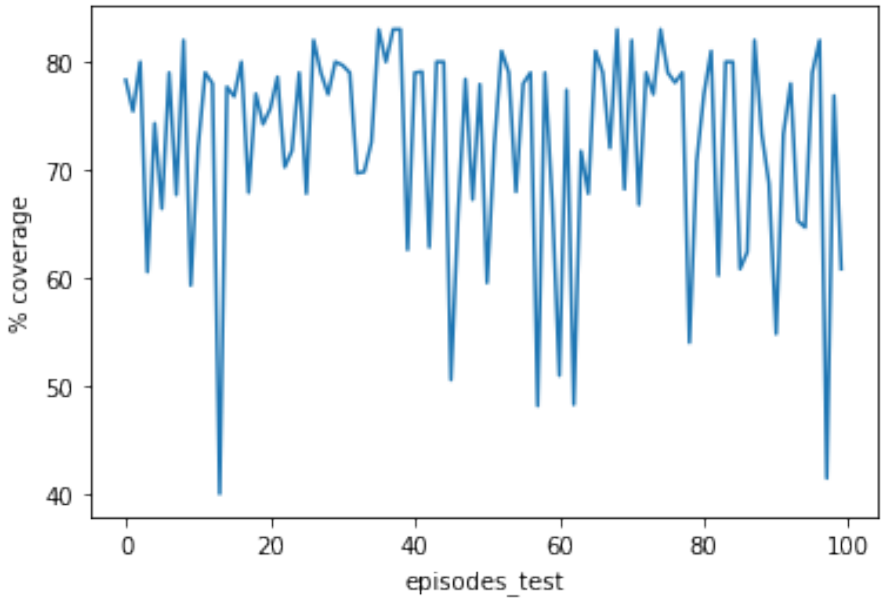

Figure 2: new model

Moving on, there is exploitation in information that it was not used previously. Specifically, when an agent is into a position he perceives the eight surrounding local areas that he can observe, in which each area can have 3 possible states unexplored, explored and obstacle. So, for each agent in each time step we store the gps coordinates for each unexplored area that he perceives but not visiting to a list of undiscovered places and we remove it (from the list) in case of visiting that position on the future. This list is being shared amongst the agents. For that purpose, a second neural network is being introduced, that its purpose is to produce actions for the agent to go to a desired destination (unvisited area based on the list above),while moving on the environment based on the local observations and on the same time avoiding obstacles. The input of this neural network is again the local observations, plus the current position and the desired position. The structure of the net is similar to the previous one, tho it has less hidden layers. Specifically, it has 4 hidden layers and the output has changed to four possible actions, that is removing the action of staying still. The net was trained again with the Reinforcement Learning algorithm (DDQN) with a reward of:

$$R = \begin{cases} -|(x_{target} - x_{current})| - |(y_{target} - y_{current})| - 2 * wall_{hit}, & \text{if position is not reached.} \\ 1000, & \text{otherwise} \end{cases}$$

After the training the neural net had 100% accuracy in both reaching its target and avoiding obstacles. When, the second neural net is enabled (via switcher neural network) it calculates the closest (undiscovered) position from its current position from the saved list of all the undiscovered places and then it produces the necessary actions until the agents reaches the target destination. In the following graphs the first one (**3**) shows on the x-axis the episodes testing and on the y-axis the binary value of finding its target position (1 went to target position and 0 did not) and the second graph (**4**) again is the wall hit for the episodes testing and again its 0.

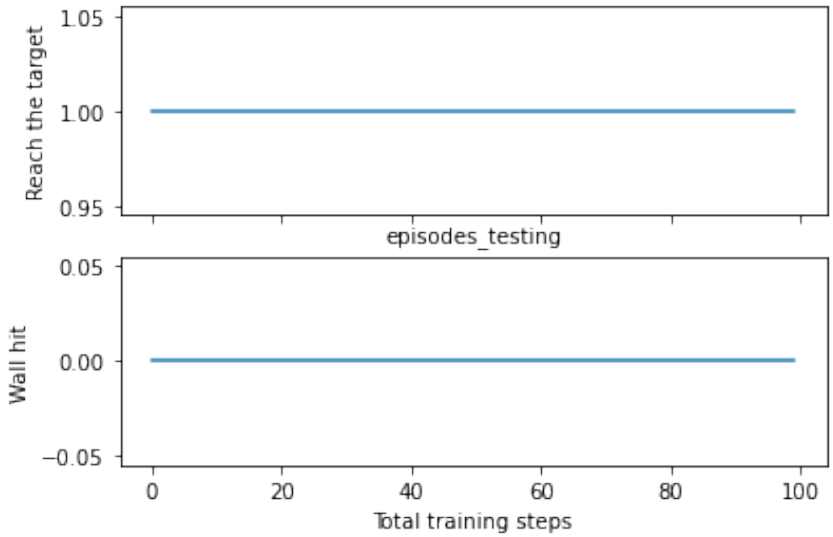

Figure 3: Neural Net for target position

Lastly, the agents need a way of choosing between the two neural networks, that are the main neural network and the secondary neural network for the closest undiscovered target position. That choice is solved with the usage of a third neural network called switcher. The switcher chooses which network will be selected in each timestep for each agent. The input of that net is again the local observations for each agent, the current and target position (closest unexplored area), plus the total number of undiscovered places and finally the total coverage for each step (total coverage for all agents). Again, that net was trained with the Double Q-learning with the reward being the total coverage gaining each timestep minus the total sum of steps, that the agent hasn't found a new area.

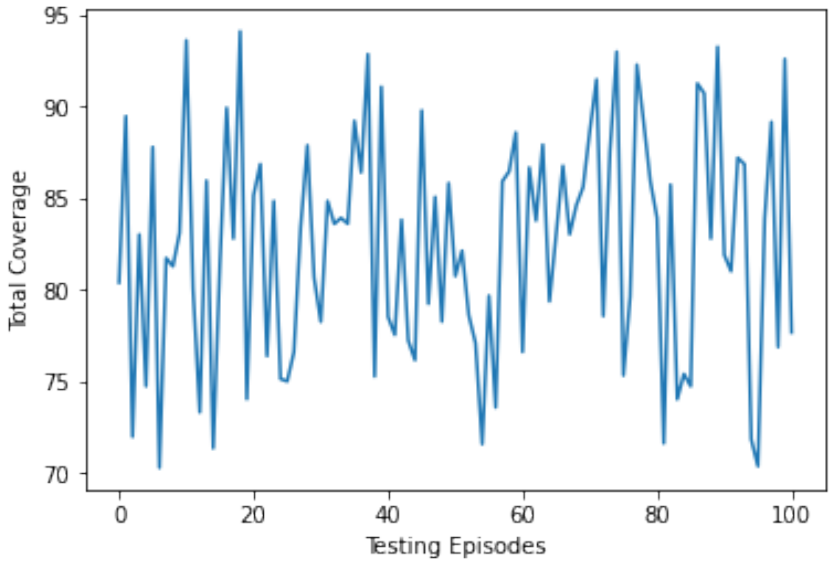

Figure 4: Model coverage with Switcher and New algorithm

3 Results

With the combination of the deeper neural net model, the algorithm for the closest target position and the switcher neural net we have a gain in total median coverage of 23% (83% from 60%). Also we see that the best episode is getting closer to the 100% coverage (95% across several testing episodes) and also the minimum is greatly increased (71%). We managed to reduce the dispersion a lot making to a more stable model. All these are summarized on the graph 4. Again, the total timestep for each agent is still 1000 individual movements.

Furthermore we show the results for four agents on 50x50 maps. The walls take roughly 10% of the total area space. The following videos has 95.8% (url-1) and 96.39 % (url-2) total coverage. In all testing cases the total wall hit number is 0 for each neural network. The change (switching) between of the two networks happens all the time but in both cases the total number of each one being used is balanced, meaning for four agents for the above case

we have 2001 total actions taken from the main neural network and 1999 actions taken from the target closest position network. But in general the total actions are around those numbers.

4 Conclusions/Future work

With the introduction of the new neural networks we achieved better results in both coverage and lower divergence between the testing episodes (meaning more stable behaviour). Furthermore, we are getting closer to the 100% total coverage with best episodes reaching at 96% coverage. Still the wall hit ratio is 0 for both neural networks.

For future work, the goals are to furthermore reduce the divergence between episodes and increase the coverage until we reach it to the 100%. Also, computational time will be added and the code will be optimized. Finally, ROS simulation hopefully will be implemented as the model will be getting optimized.

url: Path plan video

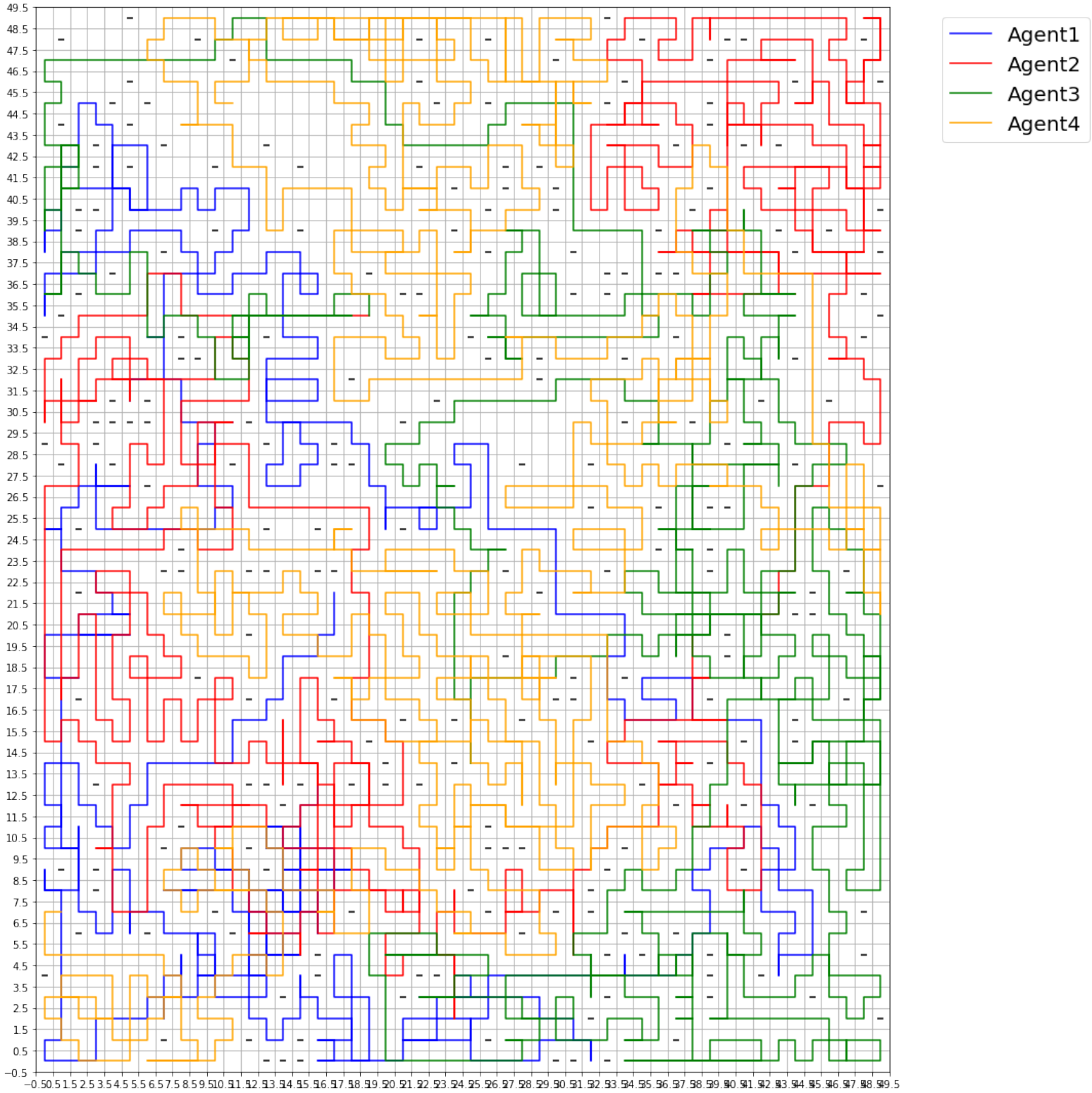

Figure 5: Four agents path, with 95.8% total coverage.

url: Path plan video

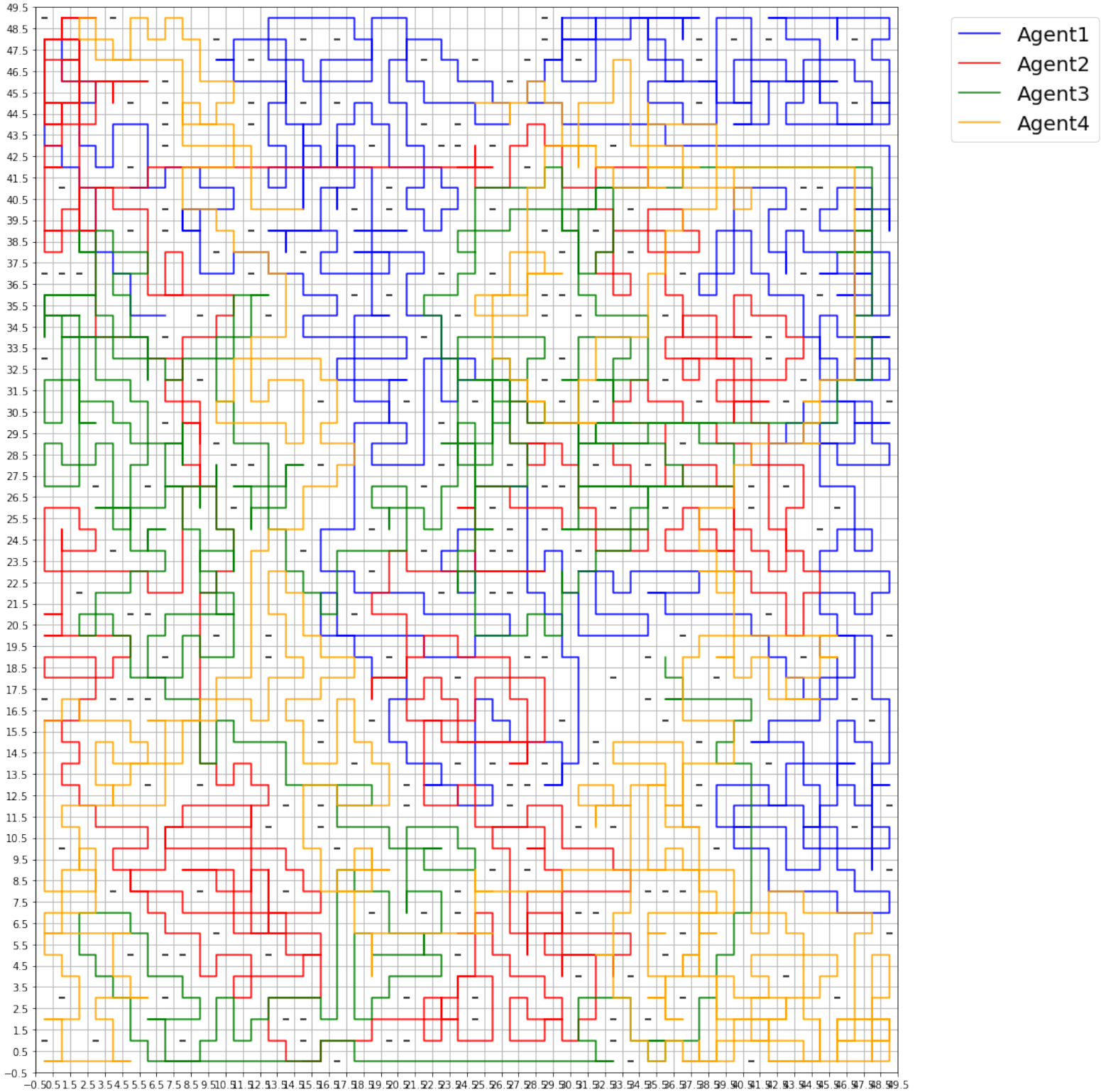

Figure 6: Four agents path, with 96.39% total coverage.