EPSRC Engineering and Physical Sciences Research Council

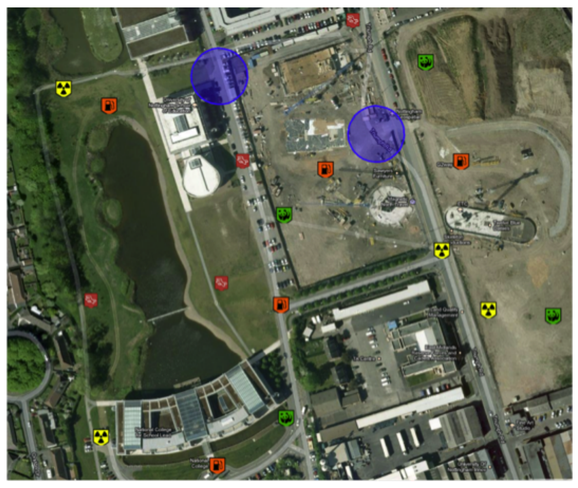
ORCHID

Multi-Agent Planning for AtomicOrchid Games

Feng Wu, Sarvapali Ramchurn, and Nicholas R. Jennings
Agents, Interaction, and Complexity Research Group
School of Electronics and Computer Science
University of Southampton

AtomicOrchid Scenario

- A dirty bomb has exploded in a park and the radioactive cloud is drifting towards hospitals, animal shelters and fuel dumps.
- Soldiers, medics, truck drivers and ambulance drivers have to form teams to evacuate the patients, animals and the fuel to safety before they are contaminated by the cloud.

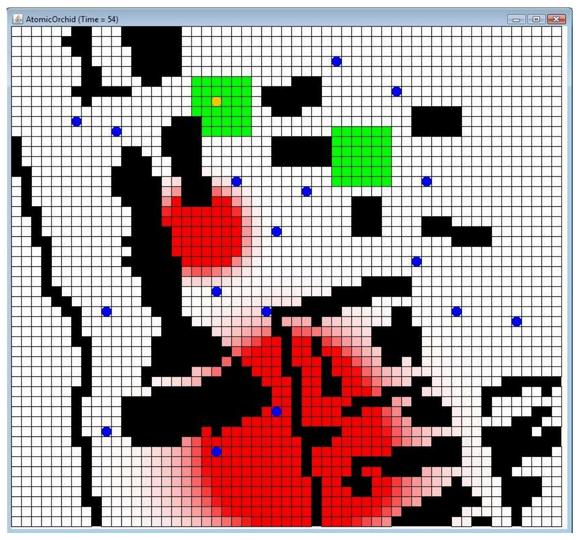


☞ The performance is evaluated depending on the value of the resources they rescue and on their radiation exposure.

Multi-Agent Markov Decision Process

The coordination problem can be modeled as a *multi-agent Markov decision process* (MMDP), $\langle I, S, \{A_i\}, P, R\rangle$, where:

- I is a set of n players.
- $S = S_r \times S_1 \times \cdots \times S_n$ are the cloud and players' states.
- A_i are player i's actions.
- P is the transition function.
- R is the reward function for completing tasks.



☞ The goal is to find a mapping from states to joint actions of all players that maximizes the long-term expected value:

$$Q(s,\vec{a}) = R(s,\vec{a}) + \gamma \sum_{s' \in S} P(s'|s,\vec{a})\, V(s')$$

where $V(s) = \max_{\vec{a}\in\vec{A}} Q(s,\vec{a})$.

Hierarchical Planning Algorithm

The overall model is huge and finding the optimal solution is computational intractable for AtomicOrchid. Thus, we approximate it by hierarchical planning with two levels:

- In the lower level, *path planning* is run for each player to find the shortest path to a task and way back to the drop-off zones from his current location given the map and the radiation cloud.
  ☞ A single-agent MDP is formulated for each player with the goal location, which can be efficiently solved by *value iteration*.
- In the higher level, *task planning* is run for the team to assign the best task to each players given the current state of all players and the radiation cloud.
  ☞ Achieving a task is treated as a macro action in this level with the assumption that every player will follow the shortest path.

Monte-Carlo Value Estimation

To assign the tasks, it is critical to know the expected value $Q(s,\vec{a})$. It does not only depend on the task reward R but also the future value V because the tasks usually cannot be completed in one shot. Thus, we use Monte-Carlo rollouts to estimate the long-term value.

- Starting from state s, the MMDP simulator simulates the movement of the players and the drifting of the radiation cloud until the termination states or a deadline has been reached.
- When selecting a task, the UCT heuristic is used to balance exploration and exploitation of the solution space.

$$\tilde{Q}(s,\vec{a}) = Q(s,\vec{a}) + c\sqrt{\frac{\ln n_s}{n_{s,\vec{a}}}}$$

Coalitional Skill Games

To form a team for a task, it is computational expensive to enumerate all possible combinations. Each player owns a skill and each task requires a set of skills. Thus, it is a coalitional skill game with Q as the characteristic function and a coalition is a team of players that can achieve the task requiring the skills.

$$Q(s,\vec{a}) = \sum_{t\in T} Q_t(s_t, a_t)$$

where T are the tasks and a_t is a coalition for task t.

- We first generate all possible coalitions for every task and then maximize the sum of the utilities with the constraint that each player can only do a task at a time.
  ☞ The optimization can be done by the max-sum algorithm.
- We also consider the feedback of the players by filtering out some of the coalitions when a player refuses to accept a pre-assigned task.
  ☞ This will change the domain space of the rejected tasks.

Conclusion and Future Work

We implemented the model and algorithms to output plans for guiding human players in the AtomicOrchid game. This has been deployed as a mobile phone app and real tests have been run in the campus of Nottingham university. In the future, we plan to:

- consider the players' speed, stamina and the terrain in the simulation of the players' movement in the lower level.
- consider the players' preference on forming a team by tracking their movement patterns in the task planning level.
- consider active sensing of the radiation cloud using UAVs.
- ..., etc.

BAE SYSTEMS

ACFR AUSTRALIAN CENTRE FOR FIELD ROBOTICS

SECURE

UNIVERSITY OF Southampton