Dr. Radu Calinescu; Dr. Daniel Kudenko. Univ. of York
Prof. Olivier Barais. Univ. of Rennes 1
Dr. Alec Banks. DSTL

Joshua Paul Riley

Safe Multi-Agent Reinforcement Learning

(Towards the engineering of trustworthy robotic teams)

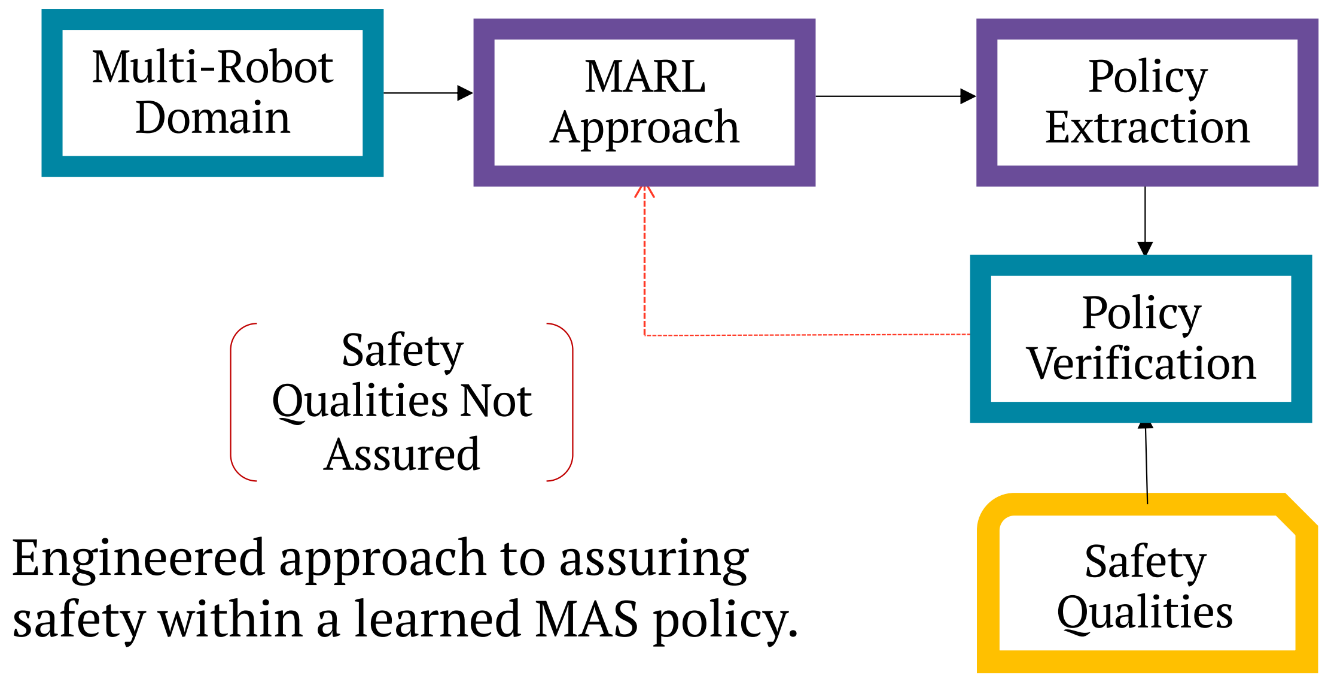
UNIVERSITY of York

1. Introduction

- Reinforcement Learning is an optimisation approach derived from behavioural psychology based on receiving information from the environment and using this information in future interactions with said environment.
- Multi-agent systems (MAS) are a collection of agents that possess some degree of autonomy, reasoning ability, and can collaborate or at a basic form, work in the same environment as other autonomous agents.
- Our view on safety is the ability to perform a task as a collaborative system while being able to predict (to a degree) and avoid behaviour that doesn't satisfy safety requirements (For example harming people or the environment).
- MARL is a very active research area with a lot of promise for progressing intelligent agents such as autonomous robotics.
- Safety within MARL is a fairly new research area with great research potential and many open problems. These open problems limit exposure of MAS to the real world.

2. Project Aims

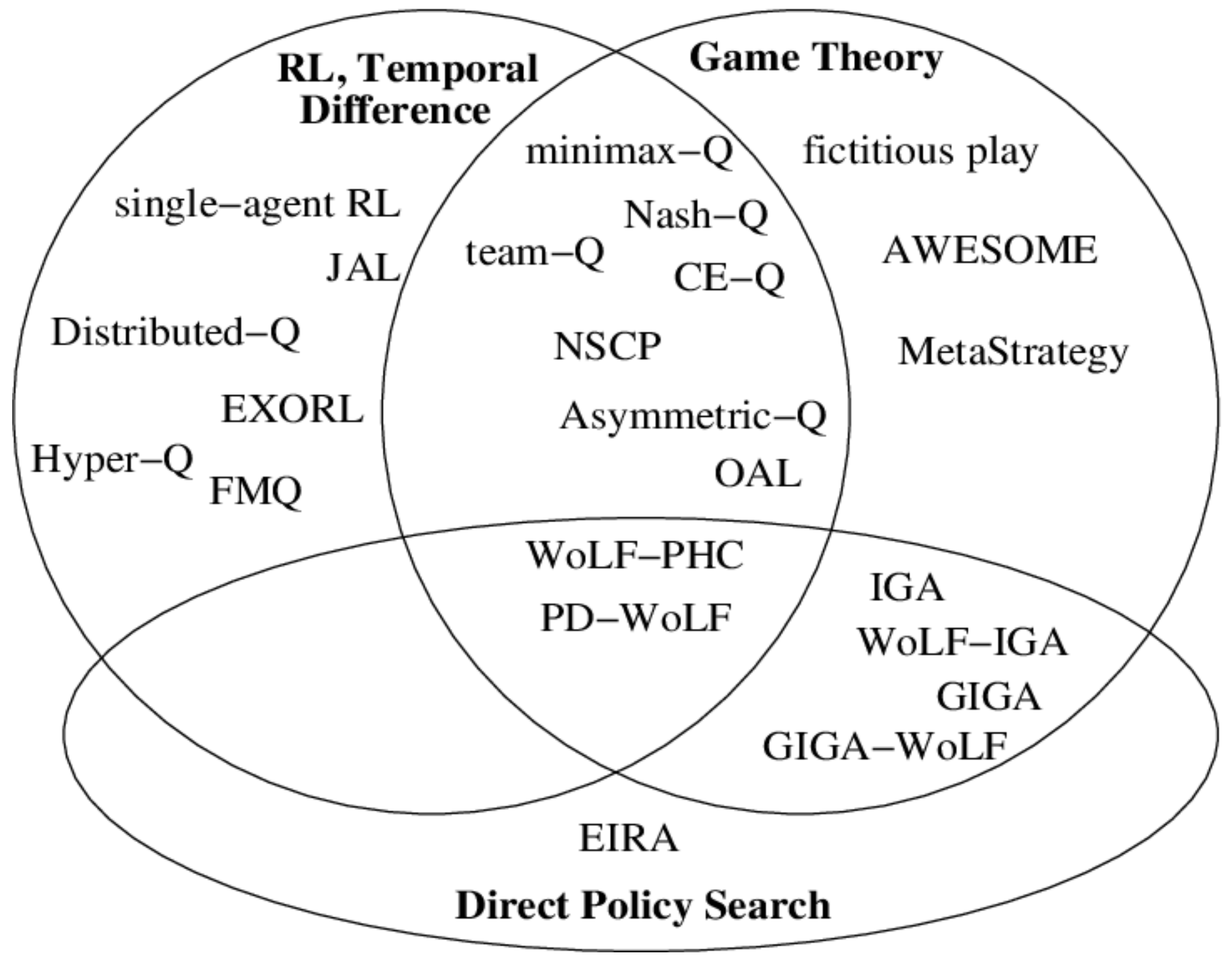
- We aim to build on reliable and well known MARL algorithms to allow a continuation of research and promote the use of these algorithms in the real world.
- We will do this using an approach which is novel to these MARL algorithms, towards securing safety in these algorithms in broad domains (Environments / Problems).
- Safety qualities should be assured for produced MARL policies.



- Engineered approach to assuring safety within a learned MAS policy.

3. MARL Methods

- Multi-agent reinforcement learning (MARL) is the process of having multiple agents work together in a shared environment (domain), as MAS was described previously, but while also learning how to best complete a goal and interact with each other.
- There are two distinct branches of research forming here, one focusing largely on neural networks and autonomous vehicles, and one focusing on primarily traditional forms of learning and more general problems.
- We take on a more traditional approach that builds on work for single-agent reinforcement learning, and we incorporate many trends from MAS and MARL while using appropriate tools such as PRISM and PRISM Games, which are briefly discussed in section 4.
- Within traditional methods exist three main types of popular MARL algorithm, as seen below, from these, we hope to build to promote safety



Buşoniu, L., Babuška, R. and De Schutter, B., (2010).

- All approaches have limitations, and it is probable that a mixture of techniques will be needed within an algorithm.
- Value iteration (Temporal Difference) approaches are questioned in practicality in comparison to Policy Iteration (Direct Policy). Still, Value Iteration is more accessible and is used in a plethora of studies with positive results.
- Team-Q; Nash-Q; Asymmetric –Q; OAL; WoLF-PHC; SA-Q.

4. Safety Methods

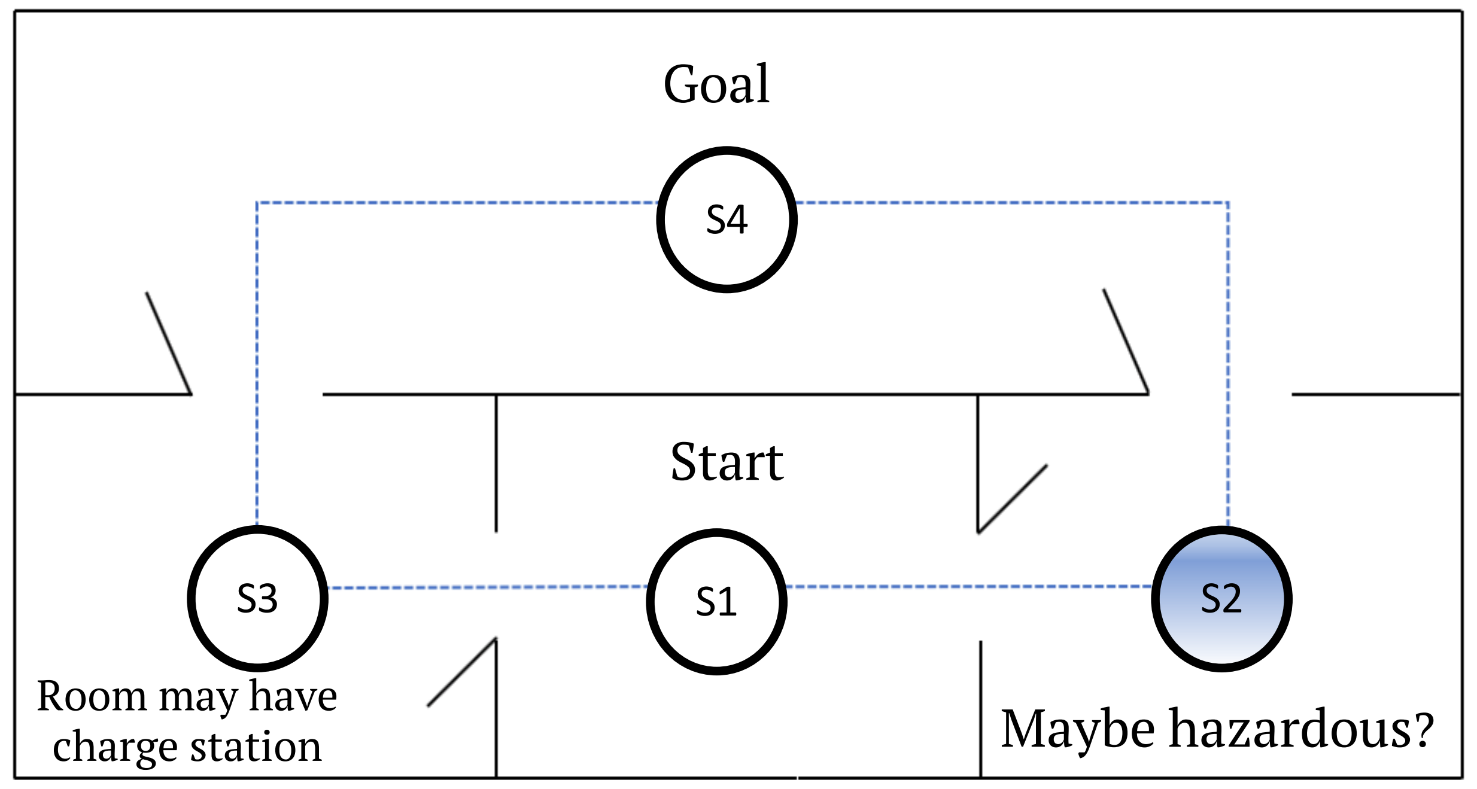
- Safety within Reinforcement Learning has prompted multiple approaches, the most prominent, however, is the behaviour constraint approach.
- Probabilistic Modelling Checking such as PRISM and are used to check for the likelihood of outcomes which can be used to choose actions to constrain.
- Using methods such as these can assure certain behaviour is followed.
- Below shows a Markov game of S1, two robots with 3 actions; The first shows the policy prior to constriction, the second after.

		Robot 1's Actions		
		Left	Right	Wait
Robot 2's Actions	Wait	5	5	0
	Right	10	10	5
	Left	-10	10	5

		Robot 1's Actions	
		Left	Wait
Robot 2's Actions	Wait	5	0
	Left	-10	5

5. Safety Concerns

- Safety concerns in MARL consist of a number of things and depend greatly on the environment (Dangerous Terrain, etc.)
- Examples (All of which can lead to unsafe behaviour)
  - Deadlock
  - Battery Life
  - Time Constraints
  - Damage to sensitive equipment
  - Damage to the robots themselves



State mapping of a floor plan (simple domain)

6. Open Problems and Assumptions

- As stated prior, MARL and Safe MARL have many open problems that require further investigation and hold promising research trajectories.

❑ State Explosion (Bellman's curse of Dimensionality).
  - Large domains (Environments).
  - System size (Amount of Agents).

❑ Fast and reliable convergence to optimal Nash Equilibrium.
  - Some algorithms have very restricting limitations.
  - Converging to optimal Nash is not always assured.
  - Converging to any Nash may take a long time.

❑ Safety qualities greatly affect behaviour.
  - Possibly optimal strategies constricted.
  - Large search spaces can equal lengthy policy checks.

- Assumptions have had to be taken during the project to avoid tackling too many problems.
  - Communication (if used) is secure.
  - Time for extensive model checking and policy creation is available.

7. Conclusion

- MARL algorithms have been adapted and tested against each other over the past two decades; this allows credible approaches to be scrutinised and adapted for safety purposes.
- A new approach to safe MARL can be taken to further the use of MARL in real-world scenarios such as search and rescue. Making use of a framework that includes credible algorithms and model checking.

Multi-drone system flying into a potentially unsafe environment (Search and Rescue)
Robotics and Perception Group, ETH Zürich. http://rpg.ifi.uzh.ch/