Safe and Robust Reinforcement Learning for Autonomous Distributed Cyber-Physical Systems

Nathaniel Hamilton

Reinforcement Learning (RL) algorithms are rapidly gaining popularity in the control and artificial intelligence research communities because they consistently show an ability to produce optimal results even when trained with inaccurate or incomplete models [1]. Instead of hard-coding, agents are programmed via reward and punishment without needing to specify how the task is to be achieved. Because of this, RL techniques are extremely attractive for robotics applications, which have complex dynamics and environments that are difficult to model. However, the behavior of these RL systems are often difficult to interpret and predict. A simple example of this can be seen in teaching hexapod robots to walk using RL. During unbounded exploration, some applied policies can cause legs to cross, get caught on each other, and pull apart, causing one or more legs to break. Similar scenarios and situations can occur in any system trained using RL. This is unacceptable for safety-critical systems where unpredictable and unsafe behavior could be the difference between life and death.

Dangerous scenarios and situations only increase with the addition of more agents into the system. As a result, multi-agent learning systems have been confined to the world of academic research [2]. Since multi-agent learning systems have potential applications in exploration, surveillance, search and rescue, intrusion tracking, and inspection, providing safety guarantees that are robust and sound could drastically change how autonomous systems are used in both military and civilian applications.

1. Statement of Research Problem

To mitigate the risks and potential hazards associated with using RL to control autonomous distributed cyber-physical systems (CPS), this research proposes to develop a framework to ensure safety requirements are upheld while optimizing control of distributed CPS. The primary target for the framework will be commercial quadrotor drones because of their accessibility and popularity in related literature. Additionally, the framework should support human control replacing the autonomous learning control. This would expand its usability to a similar case where the control of the system is unpredictable. In order to solve this problem, the approach is divided into three main tasks:

- **Task 1. Develop a Framework for Single-Agent Systems:** To develop a safe and robust RL framework for use with a distributed, multi-agent system, there first needs to be a framework successfully proven for use with one agent.
- **Task 2. Expand the Framework to Distributed, Multi-Agent Systems:** To expand the framework to include distributed multi-agent learning, a protocol for collision avoidance will be developed. Furthermore, research will be conducted into how best to apply **transfer learning** in multi-agent systems to improve performance while maintaining safety requirements.
- **Task 3. Evaluate Framework Using Commercial Quadrotors:** To evaluate the safe and robust RL framework, it will be employed on a number of commercial quadrotors and tested in various experiments controlled autonomously as well as by **human operators**.

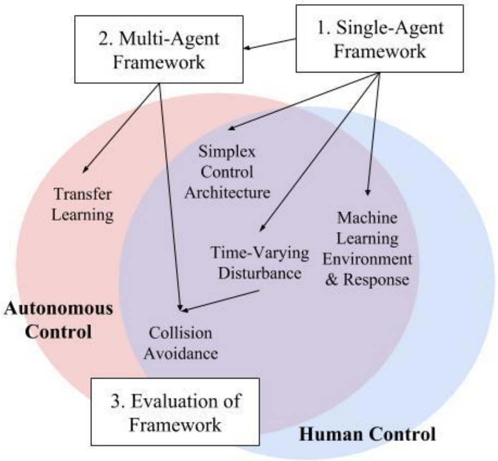
2. Multi-Agent Framework
1. Single-Agent Framework
Simplex Control Architecture
Transfer Learning
Machine Learning Environment & Response
Time-Varying Disturbance
Autonomous Control
Collision Avoidance
3. Evaluation of Framework
Human Control

Figure 1: High-level overview of the proposed procedure and major tasks.

2. Background

In order for this research to be successful, access to the following is necessary: quadrotor drones, an accurate localization system, and mentors with proven experience in formal verification and provably safe autonomy. The Verification and Validation for Intelligent and Trustworthy Autonomy Laboratory (VeriVITAL) provides access to all of these necessities and more. VeriVITAL currently owns more than 100 quadrotor drones, as shown in

Figure 2: 61 of VeriVITAL's more than 100 quadrotor drones.

Figure 2, with plans to acquisition more. The lab space has set up a small-scale indoor localization system as described in Chapter 3 of [3] using downward-facing Microsoft Kinect cameras to track drone locations in 3D space. While the system is limited to a small, confined space, work is being done to expand its usable range similar to the method proposed in [4]. In addition to these necessary items, the mentors in VeriVITAL are known for publishing notable works in the fields of Neural-Network Verification [5], Hybrid Systems Verification [6], Real-Time Reachability [7], and Safe Swarm Behavior [8]. Furthermore, VeriVITAL is involved in the **DARPA Assured Autonomy** project focusing on Assurance-Based Learning-Enabled Cyber-Physical Systems (ALC). Therefore, VeriVITAL can provide everything necessary to make this research successful.

3. Procedure

This research will develop a framework for safe and robust autonomous learning and human control of single- and multi-agent systems implemented on commercial quadrotor drones. The research goal will be achieved through the following tasks.

3.a. Task 1: Develop a Framework for Single-Agent Systems

The first research task is to develop a framework for safe and robust RL for use with a single learning agent. This will be done using a simplex control architecture utilizing the real-time reachability tool developed in [7], *rtreach*, to determine when control needs to be switched from the arbitrary RL algorithm to the safety controller. It is important that the framework works with any arbitrary RL algorithm because different RL algorithms are more effective in certain situations than others and because limiting the framework to handling only a few algorithms would be detrimental to its application.

In addition to the simplex control architecture, an offline Bayesian Machine Learning algorithm will use system response data collected from the agent to update and refine the environment and plant models used by *rtreach* to compute the reach set. Not only does this improve the accuracy of the computed reach set, but it also allows for the system to account for changes in the environment. For example, if the framework is used on a quadrotor that flies both indoors and outdoors, the reach set computed in both locations should be different, but if the models are kept the same, the computed reach set in one environment or the other will not be as accurate leading to a reduction or nullification of the computed safety guarantees.

This framework is similar to the one developed in [1] except they use Hamilton-Jacobi (HJ) reachability analysis to determine when control needs to be switched to the safety controller. However, HJ reachability is not scalable, nor is it as accurate as the face lifting method used in *rtreach*. Regardless, the experiments they use to test and validate their framework will be used for comparison and proof of concept. In addition to testing the new framework with the same experiments, the research for this task is planned to go further than guaranteeing safety when a constant unmodeled disturbance is encountered. Additional experiments would investigate use of a time-varying unmodeled disturbance like an oscillating fan so that the disturbance is present only some of the time. The main goal of this experiment would be to find a way for the learning agent to learn the pattern of the disturbance and determine a way to exploit it for use in completing its desired trajectory. From there, research would then focus on other types of disturbances like time-variable drone weight and center of mass to simulate carrying a moving object. If the framework is successful handling both types of disturbances, further experimentation would observe the robustness of the system when faced with both disturbances acting on the agent at the same time. Results from these experimentations would help determine which sacrifices in efficiency need to be made to confidently ensure safety even when faced with the most adversarial combination of disturbances.

3.b. Task 2. Expand the Framework to Distributed, Multi-Agent Systems

The second research task is to expand the framework to handle distributed, multi-agent systems. The first step in accomplishing this task is determining which method of multi-agent learning is best suited to the framework and ensuring safety. The options include applying a single learner to discover joint solutions to multi-agent problems (team learning), or using multiple simultaneous learners (concurrent learning). The current belief is that concurrent learning will work best, which is why so much time will be spent developing the framework for a single agent. However, one of the major problems faced by concurrent learning is that each learner is adapting its behaviors in the context of other co-adapting learners over which it has no control [9]. This can cause learners to make false assumptions about what the co-learners will do resulting in unsafe behaviors. Nonetheless, the safety requirements should enforce predictable behavior and remove this issue.

Regardless of which multi-agent learning method is used, a method for guaranteeing two agents do not collide will need to be developed. This can be treated as a time-varying specification as long as each drone knows where its neighbors are. Since time-varying specifications were a main focus of the single-agent framework, expanding the framework to avoid collisions with other learning agents should be fairly straightforward.

The main focus of expanding the framework to distributed, multi-agent systems will center around how to handle transfer learning. Transfer learning is the process of storing knowledge gained while solving one problem and applying it to a different but related problem. In the case of the multi-agent learning system, that could involve sharing learned information about the environment with neighboring drones. After reviewing the literature, it was concluded that this has not been investigated yet. As a result, this research could be the foundational work on how to implement transfer learning in multi-agent learning systems, or determine that transfer learning should not be implemented in this context.

3.c. Task 3. Evaluate Framework Using Commercial Quadrotors

The third and final task is to evaluate the framework's ability to provide safe and robust control. Since the simplex control architecture is capable of operating as a sophisticated geofence, it is reasonable to assume that the RL control algorithm could be switched out for human control and still provide the same safety guarantees. As a result, one desirable form of evaluation providing a **qualitative measure of success** will involve a **live demonstration** where a quadrotor's controls are given to a human with the challenge "try to crash the drone" to prove how well the safety conditions are enforced. Further evaluations for qualitative success will include multi-agent systems flying in formations and possibly interacting together to accomplish tasks. This would effectively evaluate the feasibility and use of transfer learning and demonstrate the framework's ability to handle swarm formations. These experiments would take place outdoors using GPS for localization. Success outdoors would further prove the framework's ability to handle unmodeled disturbances.

Quantitative measures of success will come from simulations tracking the number of safety violations with and without the framework. Success would involve a significant reduction in the number of safety violations, hopefully being reduced to 0. Success of RL control algorithms would also be measured by tracking the number of times the safety controller is activated with success resulting from a significant decrease as time increases. Measuring the success of transfer learning quantitatively will come from the following scenario. One quadrotor's RL control algorithm has been running for a long time and the number of switches to the safety controller are relatively low. The policies learned by that agent are then transferred to a different quadrotor with no established policies. If the number of switches to the safety controller remain low, the transfer was a success. However, if the number increases dramatically, the transfer is not a success and further work should be done to improve the transfer method.

4. References

[1] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. H. Gillula and C. J. Tomlin, "A General Safety Framework for Learning-Based Control in Uncertain Robotic Systems," *CoRR*, 2017. Available: http://arxiv.org/abs/1705.01292

[2] M. Brambilla, E. Ferrante, M. Birattari and M. Dorigo, "Swarm robotics: a review from the swarm engineering perspective," *Swarm Intelligence*, pp. 1-41, 01 Mar 2013. Available: https://link.springer.com/article/10.1007/s11721-012-0075-2

[3] N. Hervey, "Localization and Control of Distributed Mobile Robots with the Microsoft Kinect and StarL," University of Texas Arlington, Arlington, TX, 2016. Available: https://rc.library.uta.edu/uta-ir/handle/10106/25882

[4] N. Hamilton and T. T. Johnson, "Architecture for An Indoor Distributed Cyber-Physical System Composed of Mobile Robots and Fog Computing Nodes," in *Safe and Secure Systems and Software Symposium (S5)*, Dayton, OH, 2017. Available: http://www.mys5.org/Proceedings/2017/Posters/2017-S5-Posters_Hamilton.pdf

[5] W. Xiang and H.-D. J. T. T. Tran, "Output Reachable Set Estimation and Verification for Multi-Layer Neural Networks," *{IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, Mar 2018. Available: http://taylortjohnson.com/research/xiang2018tnnls.pdf

[6] A. Sogokon, P. B. Jackson and T. T. Johnson, "Verifying Safety and Persistence in Hybrid Systems Using Flowpipes and Continuous Invariants," *Journal of Automated Reasoning*, 24 Nov 2018. Available: http://www.taylortjohnson.com/research/sogokon2018jar.pdf

[7] T. T. Johnson, S. Bak, M. Caccamo and L. Sha, "Real-Time Reachability for Verified Simplex Design," *ACM Transactions on Embedded Computing Systems*, vol. 15, pp. 26:1-26:27, Feb 2016. Available: http://www.taylortjohnson.com/research/johnson2016tecs.pdf

[8] T. T. Johnson and M. Sayan, "Safe Flocking in Spite of Actuator Faults using Directional Failure Detectors," *Journal of Nonlinear Systems and Applications*, vol. 2, pp. 73-95, Apr 2011. Available: http://www.taylortjohnson.com/research/johnson2011jnsa.pdf

[9] L. Panait and S. Luke, "Cooperative Multi-Agent Learning: The State of the Art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, pp. 387-434, 05 Nov 2005. Available: https://dl.acm.org/citation.cfm?id=1090753