Reinforcement Learning Approaches to Intent Inference in Autonomous Vehicles

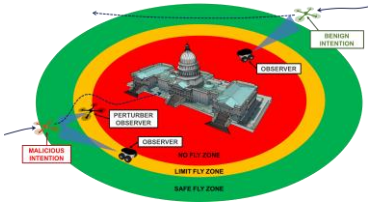
Nicola Bezzo, Mahmoud Elnaggar, Tony Lin, Peter Beling
University of Virginia

Objectives

- Develop machine learning techniques to detect malicious intent in autonomous vehicles that are compromised by cyber attacks.
- Algorithms should take as input only the actions displayed by the agent (e.g., navigation choices) and should not model the agent's internal software or communications.

Novelty of Approach

- **Inverse reinforcement learning** and sequential active learning framework in which intent is inferred from actions chosen in various states.



- Canonical scenario: UAV moving toward an unknown goal.
- Other UAVs create obstacles so we can infer intent based on response.

Deliverables

- Scenario documentation
- Vehicle demonstration, including data collection
- Inverse reinforcement learning code
- Report including conclusions and explanation of test results

Benefits to IAB

- Technical approach applicable to intent inference in CPS broadly.
- Cybersecurity of CPS is still at an early stage and is projected to grow as systems become more autonomous.
- Proposed research will contribute directly to the development of resilient techniques and increase and improve security.