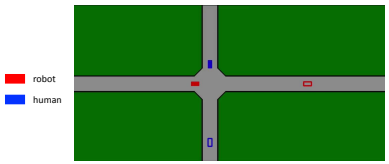
Safe Human-Interactive Control via Shielding

Jeevana Inala[1], Yecheng Jason Ma[2],
Osbert Bastani[2], Xin Zhang[3], Armando Solar-Lezama[1]

1 MIT 2 University of Pennsylvania 3 Peking University

Human-Robot Interaction



Goals:
1. Safety
2. Robot is not very conservative

(Actual) human and robot negotiate who passes first at an intersection

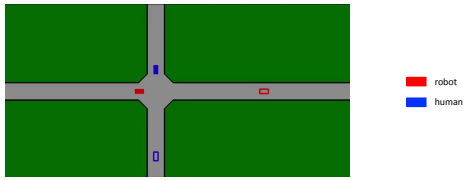
Overview

- **Key challenge:** How to model the human?
- **Naïve approaches**
  - Model human as an adversary
    - Results in conservative policies
  - Learn a model for the human
    - Model might not be accurate
- **Our approach**
  - **Human model based on a notion of fault:** We model the actions that the human and the robot are expected to take to avoid an accident
  - **Dynamics overapproximation:** We use abstract interpretation to overapproximate the reachable set and ensure that the human can always safely come to a stop

Experimental Results

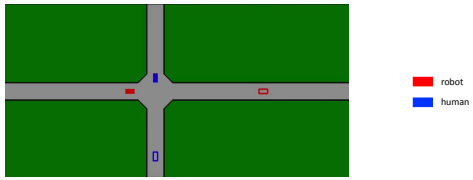
- Real humans interacting with toy self-driving simulator

I.1 Our approach is aggressive yet safe



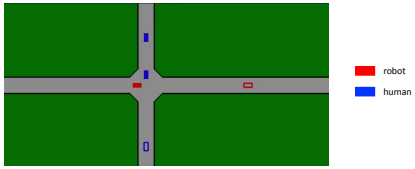
The robot ***cuts in front*** of the human leveraging the fact that a ***responsible human driver will slightly brake***

I.2 Our approach is aggressive yet safe



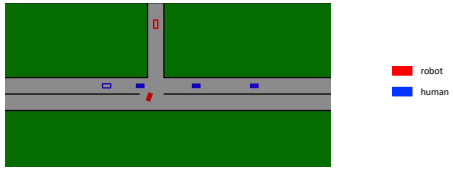
The robot ***triggers the shield to brake*** and allow the human to pass safely

II. Our approach is faster than MPC



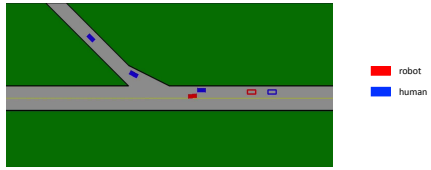
Our approach takes **less time** than an MPC approach

III. We support constraints on the robot



Here, the robot is constrained to ***not stop at the intersection***

IV. We support different backup actions



Here, the robot ***pulls over*** to the next lane as a backup policy