Uncovering the Dangers of Control Barrier Functions in Autonomous Robots

Andrew Postik, Computer Science
Mentor: Ozgur Ozmen, Assistant Professor
School of Computing and Augmented Intelligence

Introduction

Autonomous robots are increasingly being tasked with more complex, precise, and important directives, yet many rely on insecure navigation systems.

This research investigates, and hopes to mitigate, a specific vulnerability in multi-robot coordination where an adversarial agent can influence, or "herd" another robot to an undesirable location without making physical contact.

Methodology

An adversary controller can be defined with the following formulas based on the victim's state where

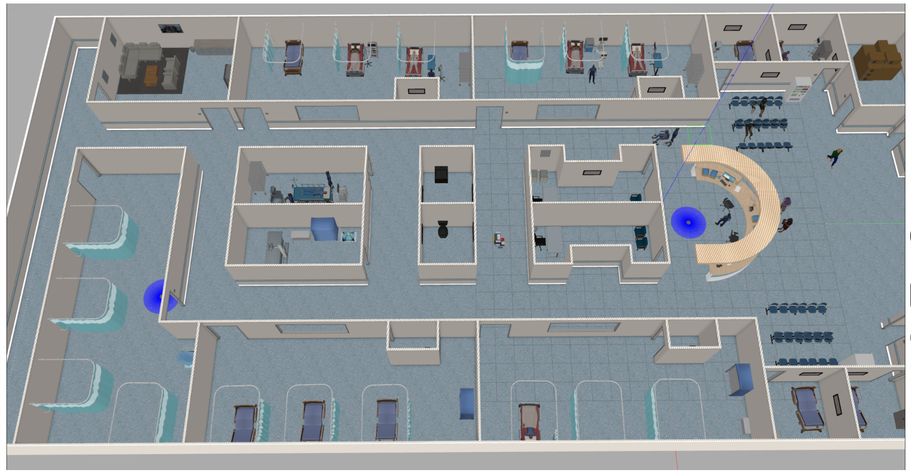
$$h = (x_r - G)^2$$
$$\dot{h} = 2(x_r - G)\dot{x_r}$$
$$\dot{x_r} = v_r$$
$$h_1 = \dot{h} + \alpha * h = 2(x_r - G)v_r + \alpha * (x_r - G)^2$$
$$\dot{h_1} = 2(x_r - G)\dot{v_r} + 2v_r^2 + 2\alpha * (x_r - G)\dot{x_r}$$
$$\dot{v_r} = a_r = K * QP(x_a, x_r)$$
$$K = \begin{bmatrix} -v sin\theta & cos\theta \\ v cos\theta & sin\theta \end{bmatrix} * g(x_r)$$
$$h_2 = \dot{h_1} + \alpha_2 * h_1$$
$$\dot{h_2} = \ddot{h_1} + \alpha_2 * \dot{h_1}$$
$$\ddot{h_1} = 2m\dot{a_r} + 2v_r a_r + 4v_r a_r + 2\alpha m a_r + 2\alpha v_r^2$$
$$m = (x_r - G)$$

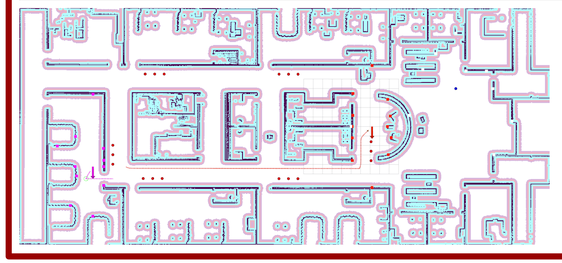
$\dot{v}_r$ is the acceleration of the adversary

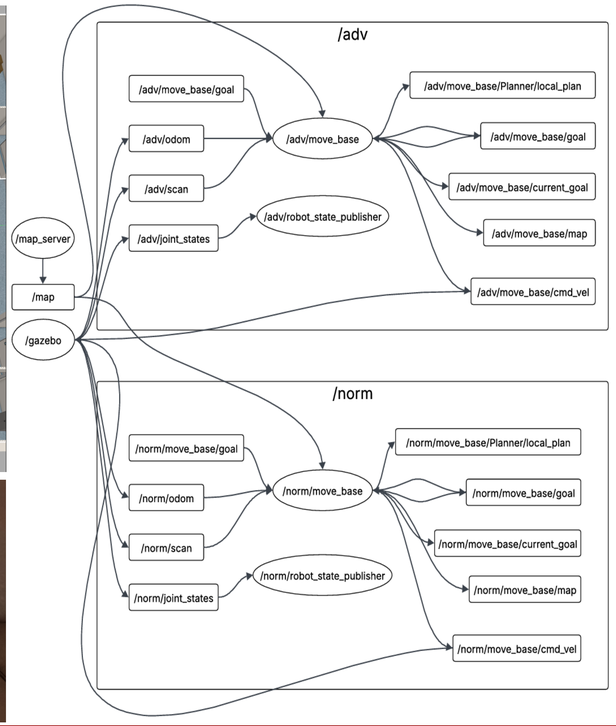
QP is the output of the victim's controller

Simulation

Using ROS2 and Gazebo, we can simulate multi-robot interactions between an adversary and a victim







Project Timeline

Set-up Simulated World and Namespaced Robots → Control Robots with a Normal and Adversary Controller → Run the Herding Attack on Varying Parameters → Observe Results and Propose Countermeasures

[1] Andrew Postik, Felipe Barreto, Hardik Parwana, Hideki Okamoto, Bardh Hoxha, Georgios Fainekos, Berkay Celik, Ozgur Ozmen. 2025. Real-Time Herding Attacks against Mobile Robots with Control Barrier Functions

FURI

Ira A. Fulton Schools of Engineering
Arizona State University