Robust Multi-Agent Reinforcement Learning for Small UAS Separation Assurance under GPS Degradation and Spoofing

Alex Zongo †, Filippos Fotiadis ★, Ufuk Topcu ★, Peng Wei †

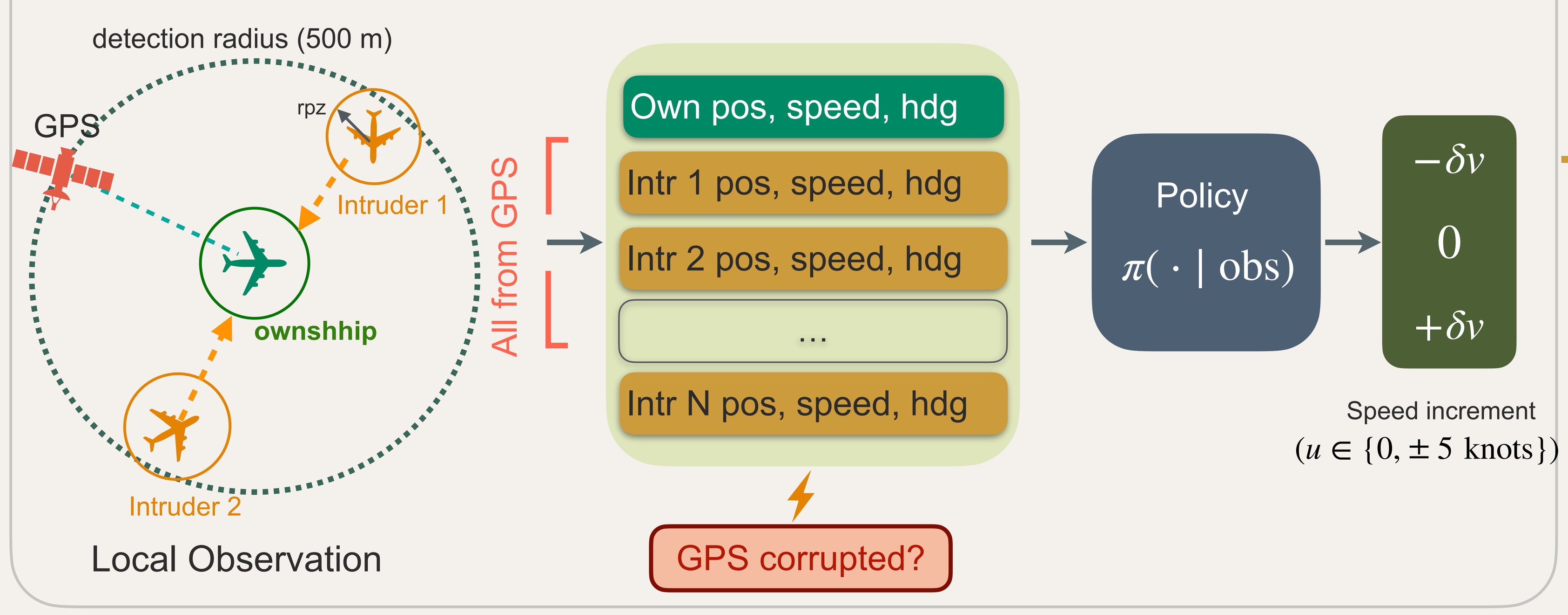
† George Washington University, Dept. Of Mechanical and Aerospace Engineering. ★ University of Texas at Austin, Oden Institute

{a.zongo, pwei}@gwu.edu {ffotiadis, utopcu}@utexas.edu

MAY 26-29 2026 ACC AMERICAN CONTROL CONFERENCE 2026 New Orleans

1 The Problem

Small UAS rely on GPS to know where they are and where nearby aircraft are.
But **GPS is fragile.** In **cooperative UAS traffic, each aircraft broadcasts its GPS position**.
When GPS is corrupted, the entire air traffic picture becomes unreliable.

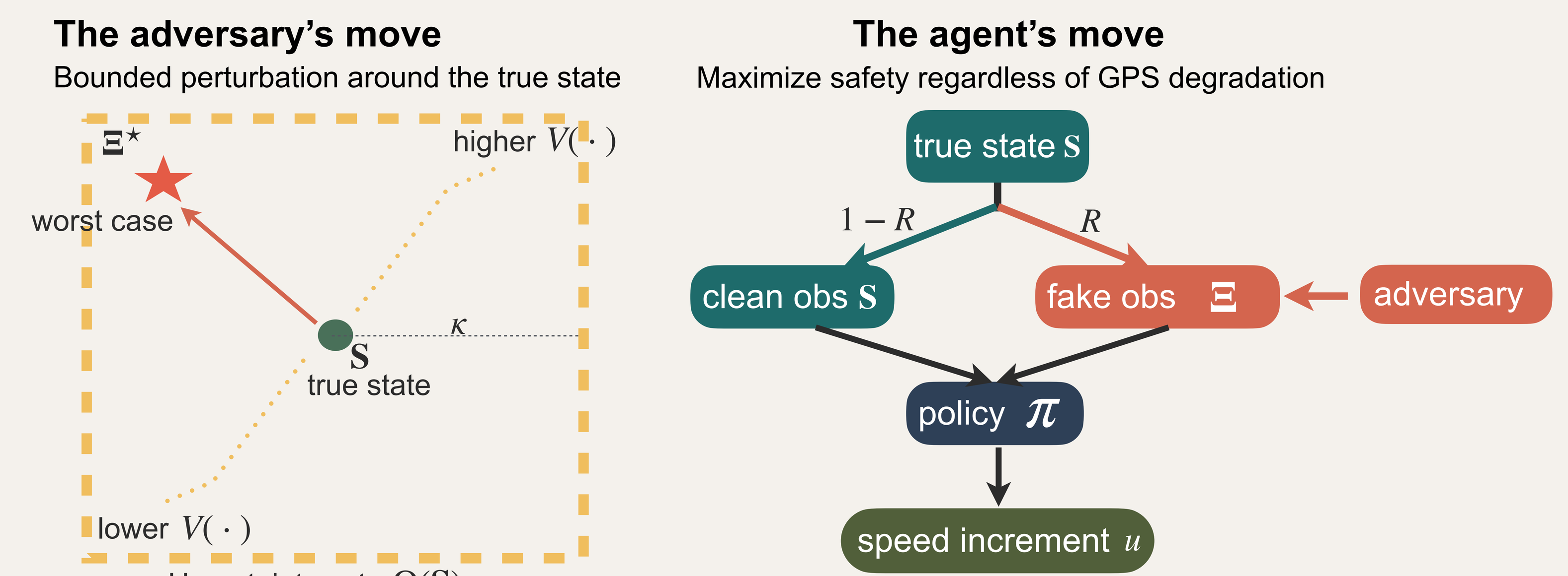


2 Adversarial Formulation

We model GPS degradation as a **two-player game**: the UAS agents try to fly safely, while an adversary corrupts their observations to cause collisions.

The adversary's move
Bounded perturbation around the true state

The agent's move
Maximize safety regardless of GPS degradation



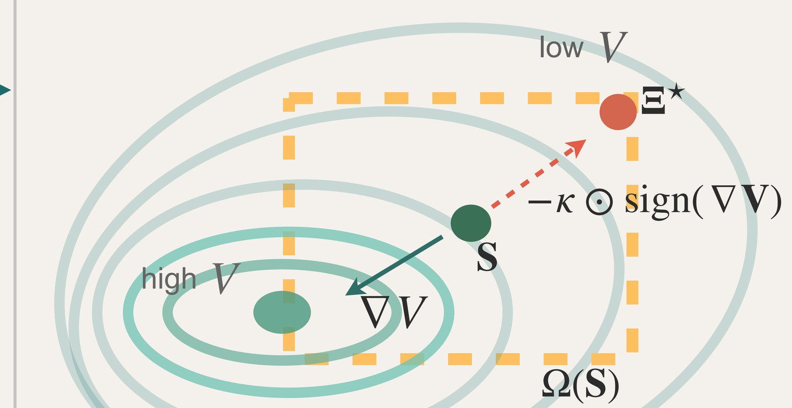
The policy must perform well under both scenarios.

$$\max_{u_t} V^{\pi}_{\text{rob}}(\mathbf{S_t}) = \max_{\mathbf{u_t}} \mathbb{E}_{\pi}\Big[\mathbf{r}(\mathbf{S_{t+1}}, \mathbf{u_t}) + \gamma\Big((\mathbf{1}-\mathbf{R})\mathbf{V}^{\pi}(\mathbf{S_{t+1}}) + \mathbf{R} \min_{\Xi \in \Omega(\mathbf{S_{t+1}})} \mathbf{V}^{\pi}(\Xi)\Big) \,\Big|\, \mathbf{S_t}\Big]$$

3 Close-form Solution

We derived a closed-form expression approximating the worst-case corrupted observation.

How works
The GPS degradation moves to the lowest value corner of the box.



The formula

$$\Xi^{\star} = \mathbf{S} - \kappa \odot \text{sign}(\nabla \mathbf{V})$$

Why this matters

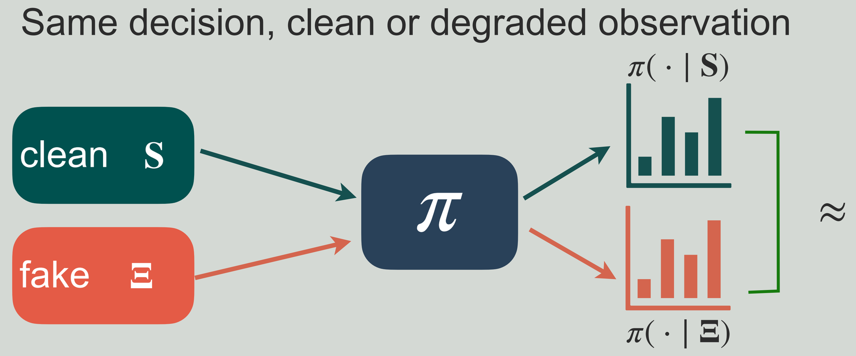
Adversarial training
Iterative, unstable, expensive

Closed-form (ours) $O(\kappa^2)$
One-shot, linear-time, exact to

4 Robust Training

Invariance regularizer

Same decision, clean or degraded observation



penalize : $\text{KL}(\pi(\cdot \mid \mathbf{S}) \,\|\, \pi(\cdot \mid \Xi))$

Anchor regularizer

Stay close to the nominal behavior (teacher policy)

1. **keeps robust policy close to nominal**
2. **prevents sacrificing clean performance**

Proposition 1: safety guarantee

If $\text{KL}(\pi(\cdot \mid \mathbf{S}) \,\|\, \pi(\cdot \mid \Xi^{\star})) \le \mathbf{B} \rightarrow$ performance loss $\le Q_{\max}\sqrt{\mathbf{2B}}$

What this means:

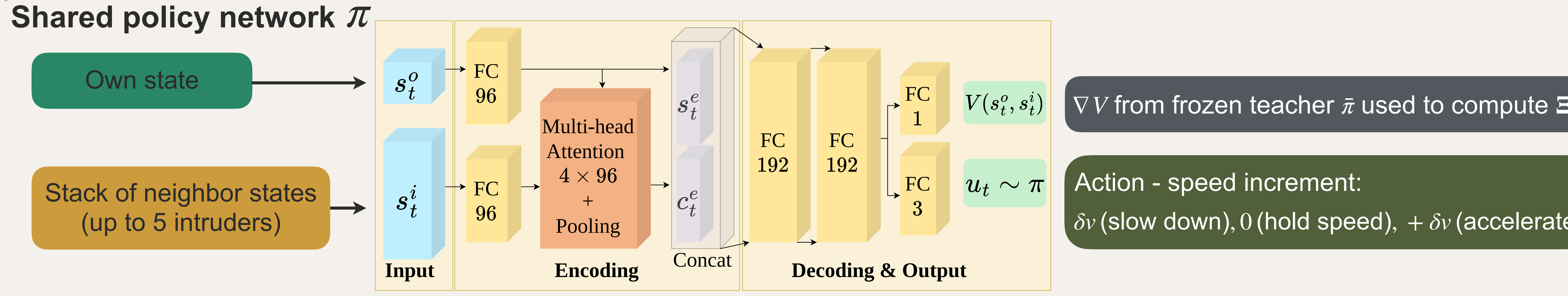
Small B reduces the loss and tightens the regularization, yielding a safer policy.

5 Training and Execution

Training (centralized)

All sUAS experiences are pooled into a single batch to update the shared policy

Shared policy network π



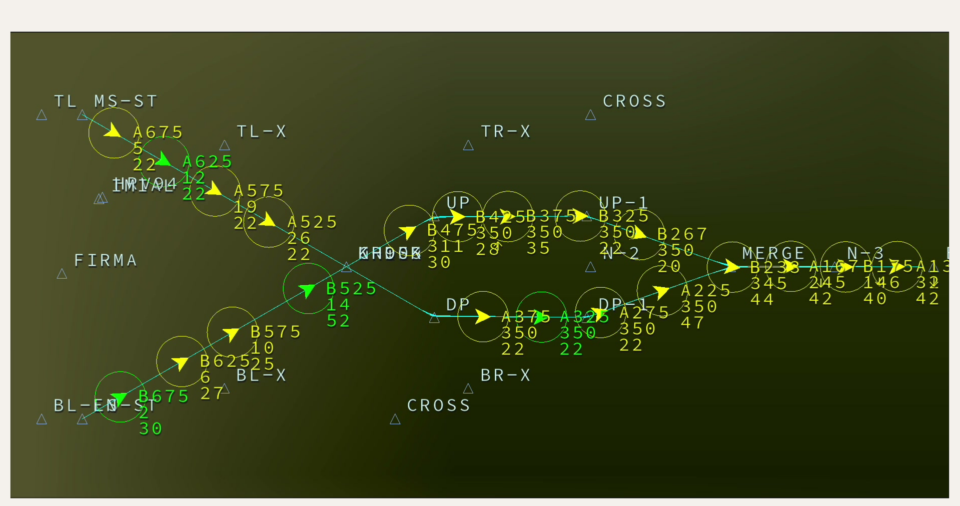
Execution (decentralized)

Each sUAS runs its own copy of π using only its local observation.

1. **Phase 1:** Nominal training (R=0) $\rightarrow \bar{\pi}$ (teacher policy)
2. **Phase 2:** Robust training (curriculum on R)

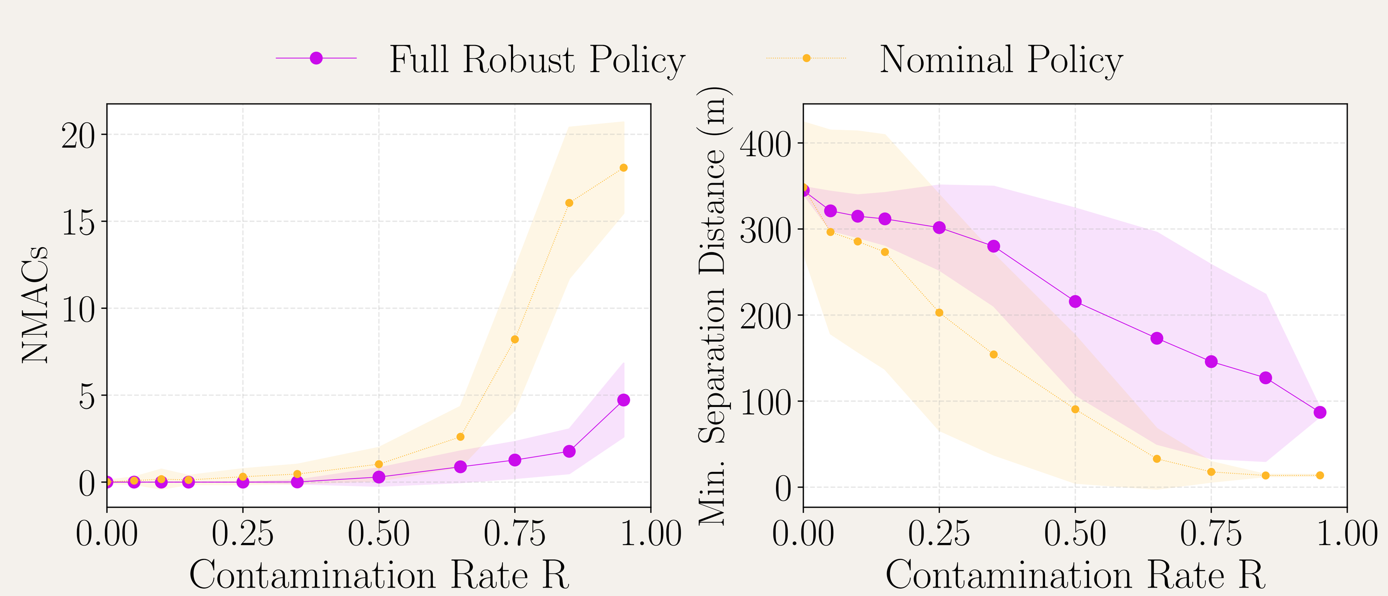
6 Results

Trained and tested on BlueSky Air Traffic Simulator with the Amazon MK30 performance parameters on an en route airspace setting for sUAS package delivery. Results are averaged over 100 episodes.



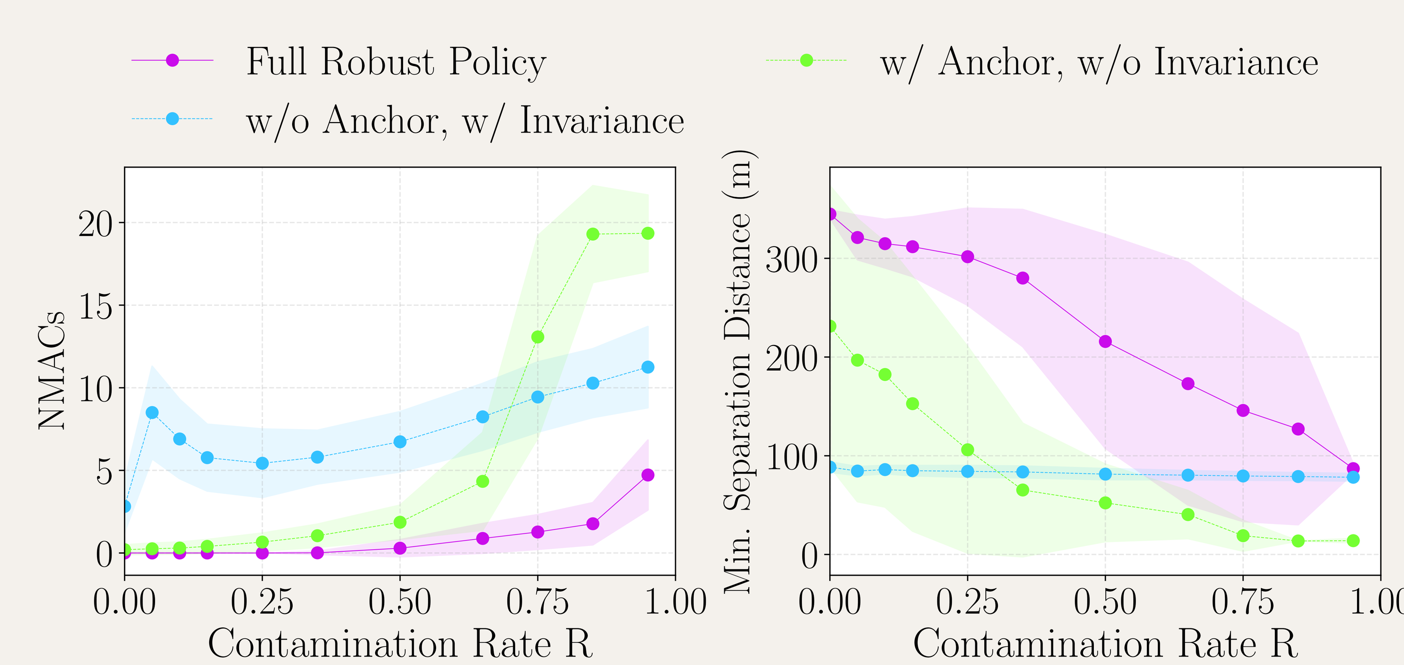
Aircraft (triangles) traverse structured routes through waypoints, with separation buffers shown as circles. Yellow indicates aircraft approaching loss of separation; green indicates nominal separation status. Labels display flight ID, speed (knots), and altitude (ft). Despite adversarial GPS perturbations, all aircraft maintain safe separation.

Safety performance under increasing observation degradation R. *Left*: Near mid-air collision (NMAC) count of the small UAS. *Right*: Minimum separation distance between the aircraft agents achieved per episode.



The robust policy maintains near-zero NMACs through R=0.35 and degrades gracefully beyond, while the nominal policy deteriorates sharply after.

Ablation study isolating the contributions of invariance and anchoring regularization. *Left*: NMAC count. *Right*: Minimum separation distance.



Invariance regularization without anchoring (blue) destabilizes training, yielding poor performance even at R=0. Anchoring alone (green) preserves nominal behavior but degrades sharply at high R. The full method (purple) combines both regularizers for consistent performance.

Take-Aways & Acknowledgement

1. A multi-agent RL approach to conflict resolution and separation assurance on en route sUAS package delivery shows promise in maintaining safety even under GPS degradation and spoofing.
2. A closed-form adversary eliminates the need for adversarial co-training while providing provable safety bounds.
3. The dual regularization, i.e., invariance plus anchoring, is essential: neither alone suffices.
4. The approach applies to any cooperative multi-agent system where sensor observations can be corrupted.

This work was supported by NASA under grant 80NSSC24M0070, and by ONR under grant N00014-22-1-2703