Secure Autonomy for Contested Environments

Miroslav Pajic

CPSL@Duke
Department of Electrical and Computer Engineering
Department of Computer Science
Department of Mechanical Engineering and Material Science
Duke University

Duke
PRATT SCHOOL of ENGINEERING

DUKE ROBOTICS

Offline Reinforcement Learning with Off-Policy Evaluation

- Off-policy evaluation (OPE) is important for **filling the gap between training offline reinforcement learning (RL) controllers and choosing which one to deploy online**

Platform Testing	Platform Not Available	Platform Not Available	Platform Not Available	Next Platform Test

Phase *I* -- *Testing*
Data collection using the latest controller.

Phase *II* – *Offline RL*
Use the updated experience dataset to fine-tune existing controllers or train new ones.

Phase *III* -- *OPE*
Estimate the controller candidates ***offline*** and select the one with best performance.

Phase *IV* -- *Testing*
Run the best performing controller selected, and collect new trajectories.

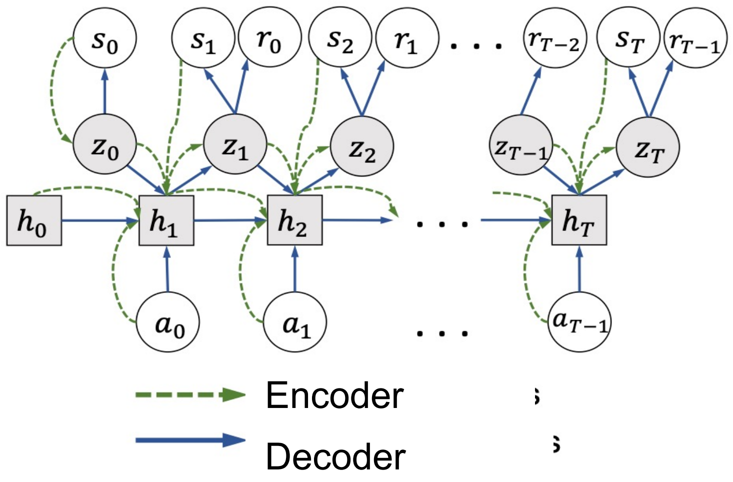
Variational Latent Branching Model (VLBM) for OPE (ICLR23)

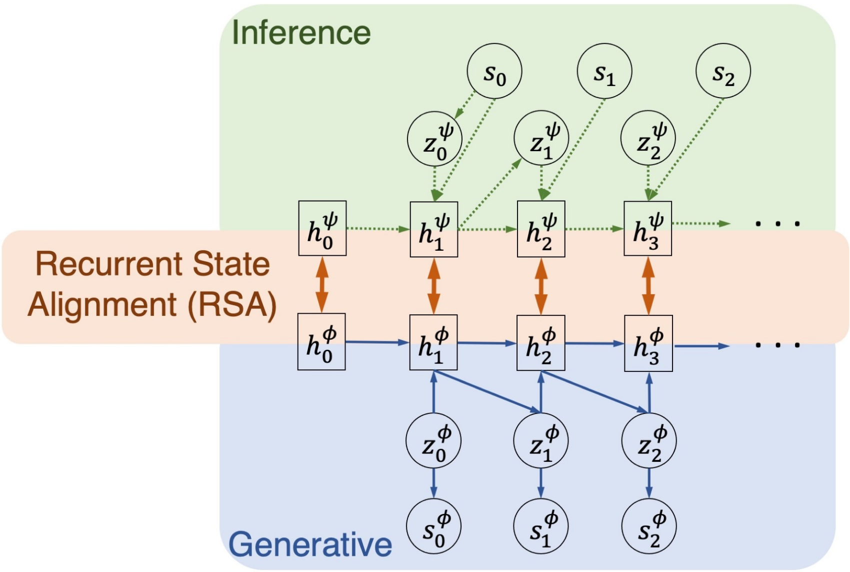
- Formulate a latent space where latent variables can transit over time $p_\phi(z_t|z_{t-1}, a_{t-1})$
- Both encoder and decoder are **LSTMs**
- The encoder infuse the knowledge of the environment into the latent space
- The decoder generates synthetic trajectories **over time**

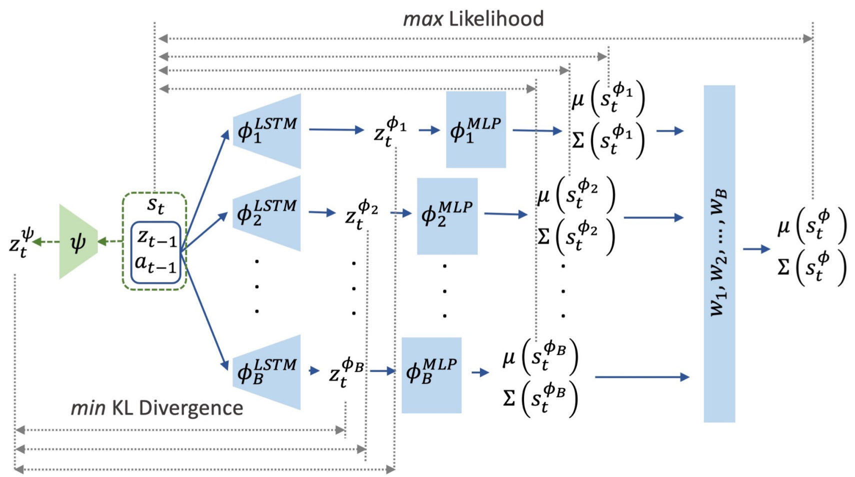
- Recurrent state alignment (RSA)
  - To mitigate the effect that decoder starts working **long after the encoder encodes the entire trajectory**
  - Minimize the **mean pairwise error** between LSTM states of encoder and decoder

- Branching for the decoder
  - Multiple decoders sample from the encoder to reduce variabilities possibly caused by, e.g., **random** initialization and **stochasticity** during training
- Overall training objective (maximize)

  ELBO – RSA + log_likelihood_for_each_branch





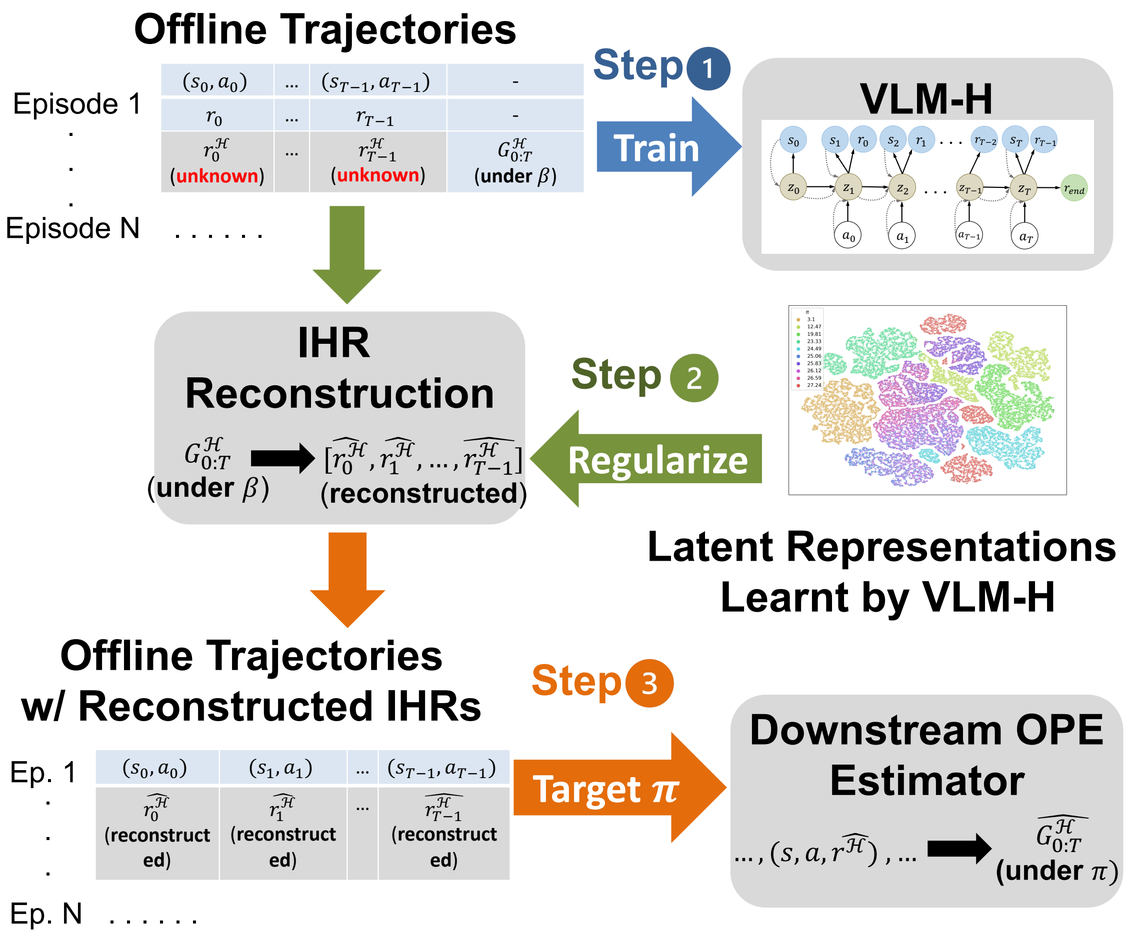


Off-Policy Evaluation for Sparse/Human Feedback (NeurIPS23)

Unknown **Immediate Human Rewards (IHRs)**

We assume that the IHRs, $r_t^{\mathcal{H}}$, are not observable. Instead, the cumulative human return, $G^{\mathcal{H}} = \sum_t \gamma^t r_t^{\mathcal{H}}$, is available **at the end of each episode (i.e., extremely sparse)**.

Objective: Given **a fixed set of offline trajectories collected by a behavioral policy** β, estimate the **expected total human return** over the unknown state-action visitation distribution ρ^π of the target (evaluation) policy π -- $\mathbb{E}_{(s,a)\sim\rho^\pi}[\sum_t \gamma^t r_t^{\mathcal{H}}]$.

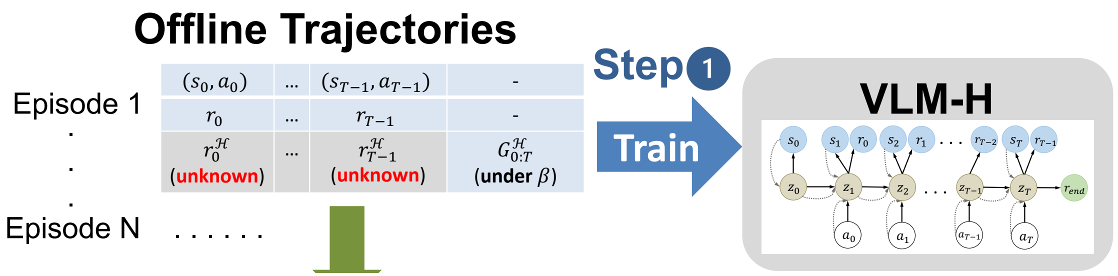


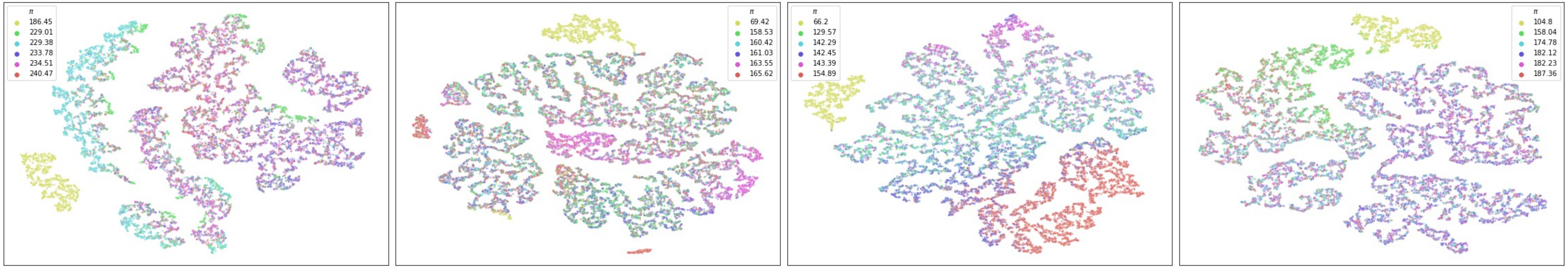
Off-Policy Evaluation for Sparse/Human Feedback (NeurIPS23)

Duke
PRATT SCHOOL of ENGINEERING

Unknown **Immediate Human Rewards (IHRs)**

We assume that the IHRs, $r_t^{\mathcal{H}}$, are not observable. Instead, the





Patient #0 Patient #1 Patient #2 Patient #3

Ol

trajectories collected by a behavioral policy β, estimate the **expected total human return** over the unknown state-action visitation distribution ρ^{π} of the target (evaluation) policy π -- $\mathbb{E}_{(s,a)\sim\rho^{\pi}}[\sum_t \gamma^t r_t^{\mathcal{H}}]$.

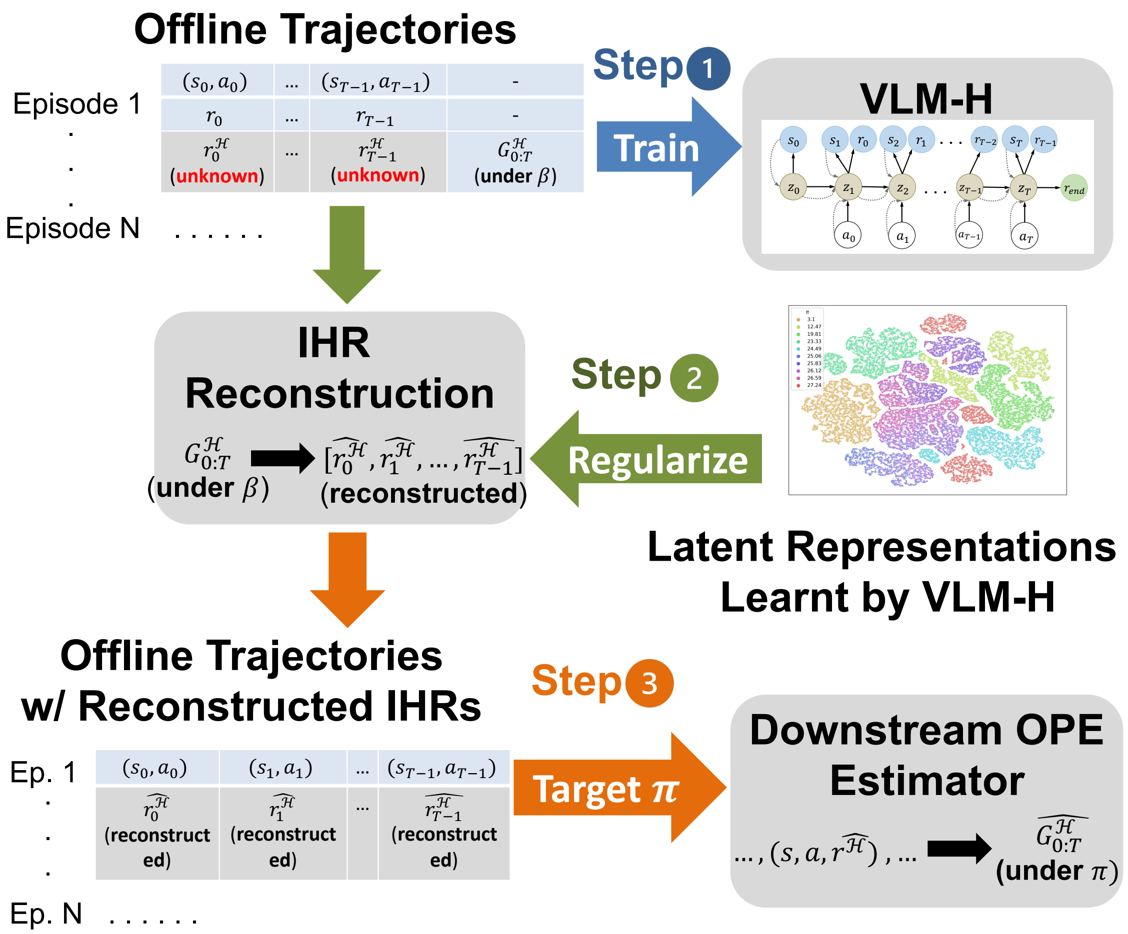
Off-Policy Evaluation for Sparse/Human Feedback (NeurIPS23)

Duke PRATT SCHOOL of ENGINEERING

Unknown **Immediate Human Rewards (IHRs)**

We assume that the IHRs, $r_t^{\mathcal{H}}$, are not observable. Instead, the cumulative human return, $G^{\mathcal{H}} = \sum_t \gamma^t r_t^{\mathcal{H}}$, is available **at the end of each episode (i.e., extremely sparse)**.

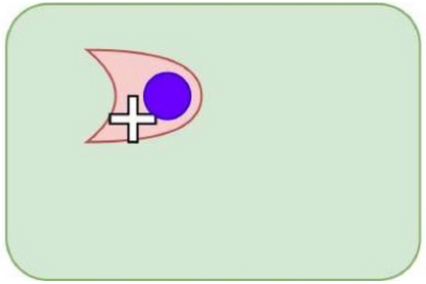
Objective: Given **a fixed set of offline trajectories collected by a behavioral policy** β, estimate the **expected total human return** over the unknown state-action visitation distribution ρ^π of the target (evaluation) policy π -- $\mathbb{E}_{(s,a)\sim\rho^\pi}[\sum_t \gamma^t r_t^{\mathcal{H}}]$.



Adversarially Robust Control & Decision Making (ICLR24*, ICRA24*)

Max-Min Optimization

$$\max_{\theta \in \Theta} \min_{\phi \in \boxed{\Phi}} R(\theta, \phi)$$

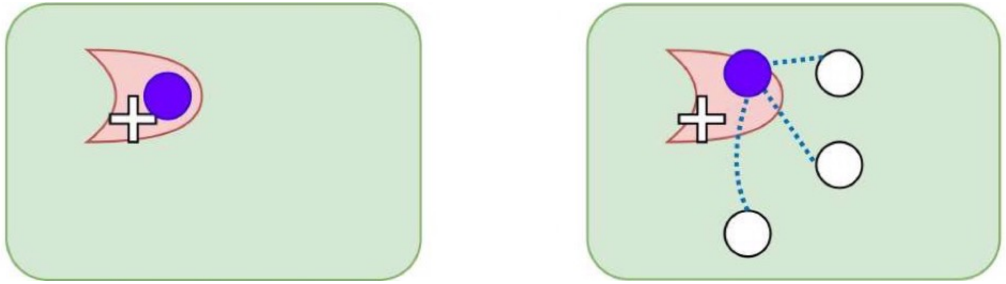


- Inner minimization problem is difficult to solve → local-optimum
- Worst-case optimization can be over-conservative for *unrealistic* adversary (i.e., overly capable)

Adversarially Robust Control & Decision Making (ICLR24*, ICRA24*)

Max-Min Optimization

$$\max_{\theta \in \Theta} \min_{\phi \in \hat{\Phi}} R(\theta, \phi) = \max_{\theta \in \Theta} \min_{\phi_1, \ldots, \phi_m \in \hat{\Phi}} \min_{\phi \in \{\phi_i\}_{i=1}^{m}} R(\theta, \phi)$$



Learners instead of fixed adversaries

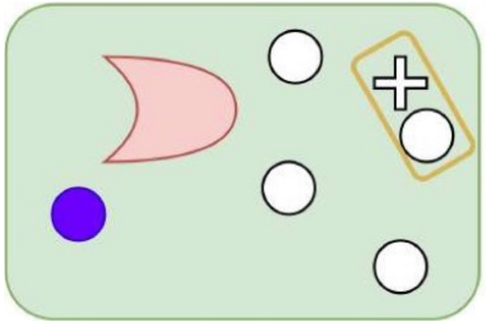
Efficient approximation of the inner optimization i.e., the size of adversary herd is upper-bounded to obtain sufficient approximation precision.

Adversarially Robust Control & Decision Making (ICLR24*, ICRA24*)

Duke PRATT SCHOOL of ENGINEERING

Max-Min Optimization with Adversarial Herd – Optimization Over Worst-*k* Adversaries

$$\max_{\theta \in \Theta} \min_{\phi \in \Phi} R(\theta, \phi) \Rightarrow \max_{\theta \in \Theta} \min_{\phi_1, \ldots, \phi_m \in \hat{\Phi}} \frac{1}{I_{\theta, \hat{\Phi}, k}} R(\theta, \phi_i)$$



Resolving Potential Over-Pessimism

i.e., modify the objective from optimizing its worst-case performance, to optimizing its average performance over the worst-k adversaries

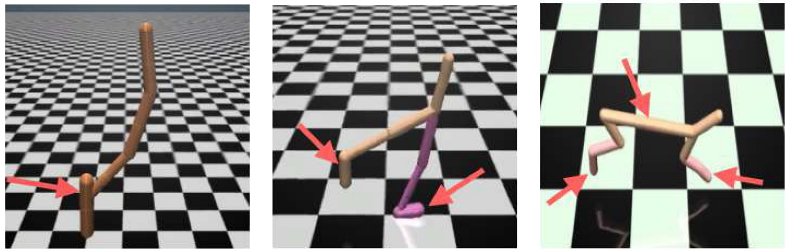
If we choose a set of adversaries that are different enough, then the number of adversaries needed to approximate the inner optimization problem is in linear order of the desired precision.

If our objective is to use adversarial herd to approximate accurately with high probability, instead of an almost sure approximation, then the number of required adversaries can be reduced.

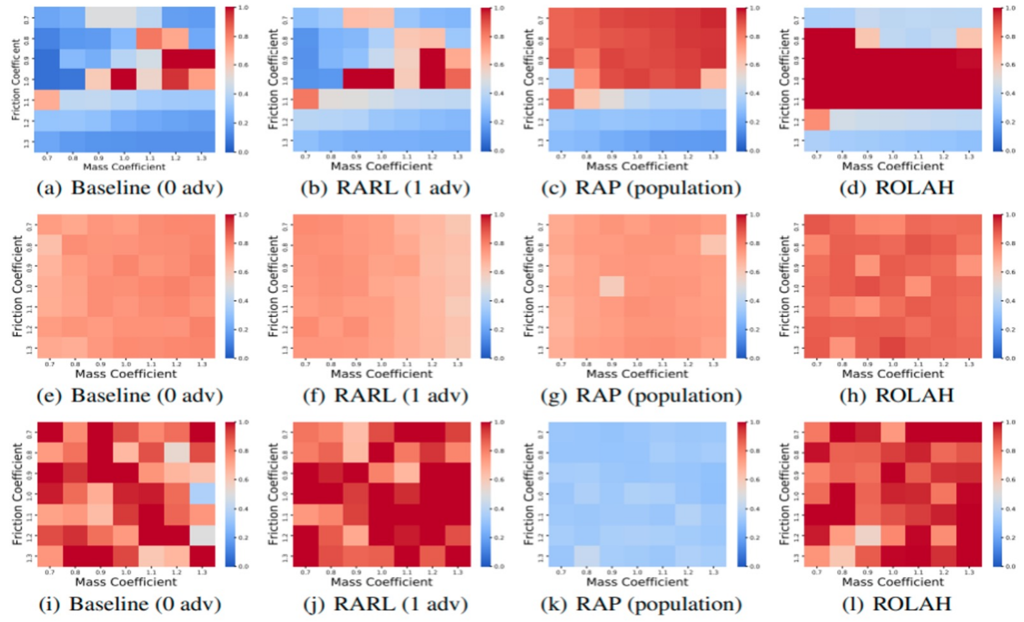
Adversarially Robust Control & Decision Making (ICLR24*, ICRA24*)

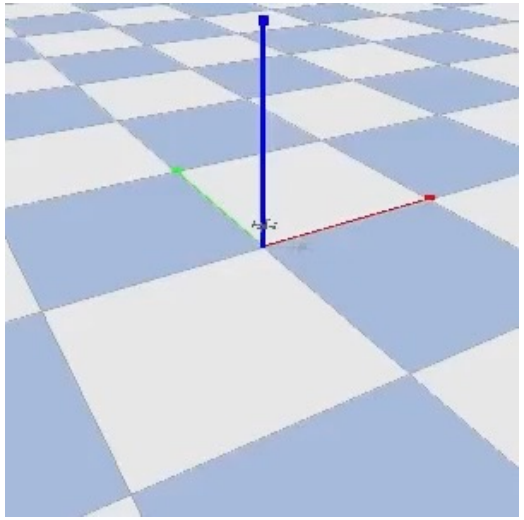
Max-Min Optimization with Adversarial Herd – Optimization Over Worst-*k* Adversaries

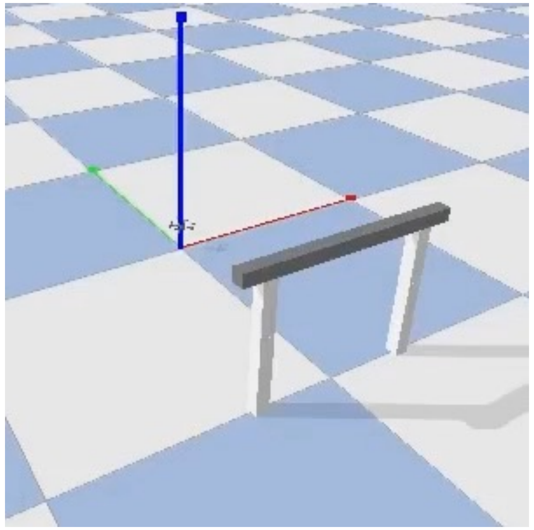
$$\max_{\theta \in \Theta} \min_{\phi \in \Phi} R(\theta,\, \phi) \;\Rightarrow\; \max_{\theta \in \Theta} \min_{\phi_1,\, \dots,\phi_m \in \hat{\Phi}} \frac{1}{I_{\theta,\, \hat{\Phi},k}} R(\theta,\, \phi_i)$$



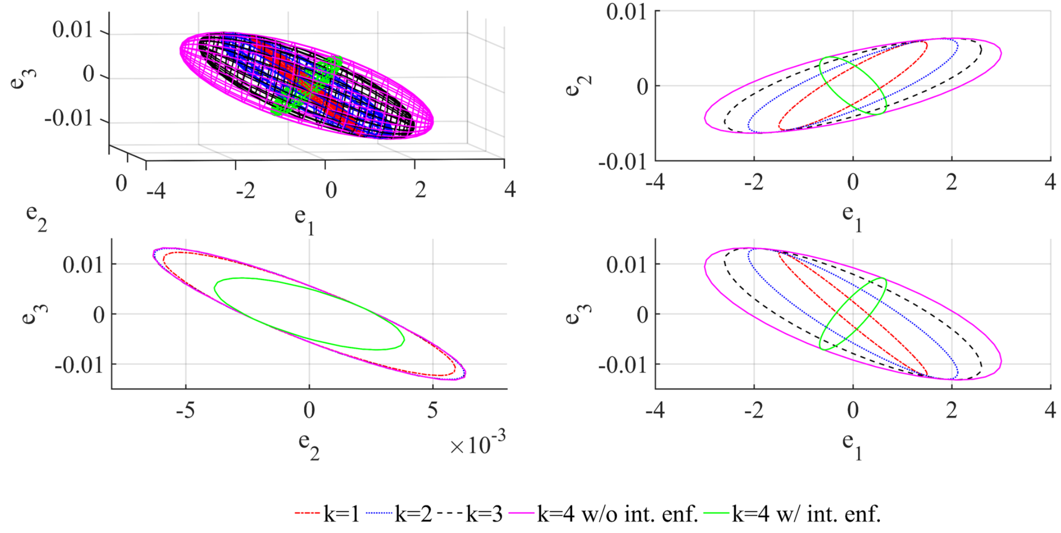
(a) Hopper (b) Walker2d (c) Half-Cheetah







How can we analyze the impact of different attack vectors on CPS (i.e., QoC)?



General Vulnerability Analysis for False Data-Injection Attacks on Control Systems (TAC24a*, TAC24b)

Duke
PRATT SCHOOL of ENGINEERING

An attack sequence

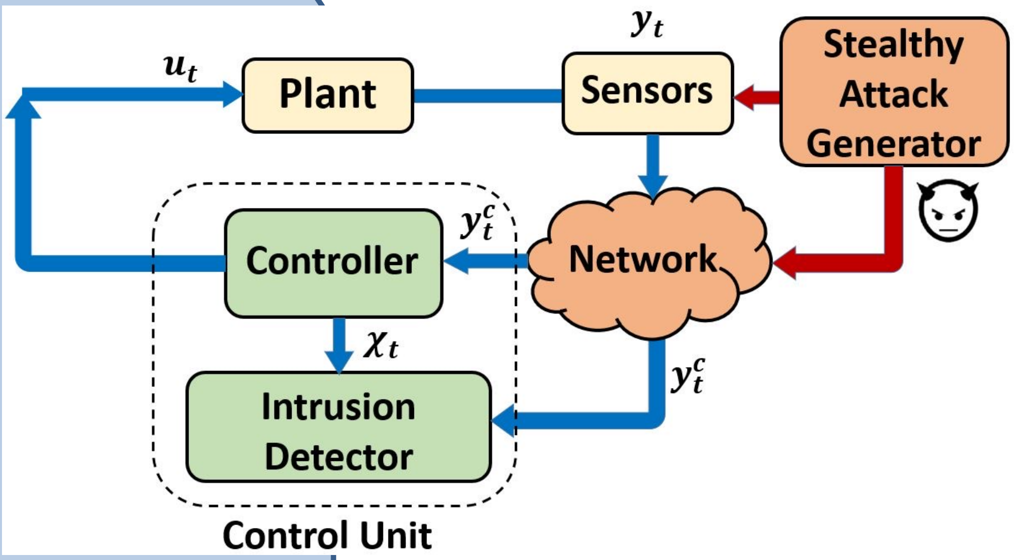
- is **strictly stealthy** iff

$$\boldsymbol{KL}\Big(\boldsymbol{Q}(Y_{-\infty}^{-1}, Y_0^a : Y_t^a)||\boldsymbol{P}(Y_{-\infty} : Y_t)\Big) = 0$$

for any $t \geq 0$,

- is **ϵ-stealthy** if

$$\boldsymbol{KL}\Big(\boldsymbol{Q}(Y_{-\infty}^{-1}, Y_0^a : Y_t^a)||\boldsymbol{P}(Y_{-\infty} : Y_t)\Big) \leq \log(\frac{1}{1-\epsilon^2}) \text{ for any } t \geq 0.$$



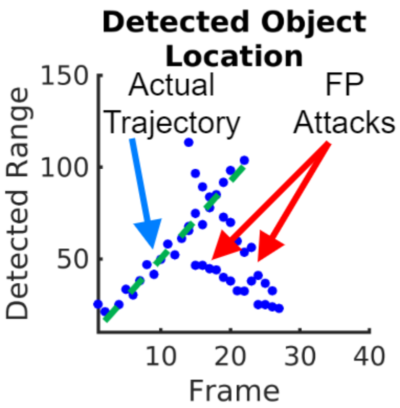
The system is (ϵ, α)-attackable for arbitrarily large α and arbitrarily small ϵ, if the closed-loop dynamics is incrementally exponentially stable (IES) in the set S and the open loop dynamics is incrementally unstable in the set S.

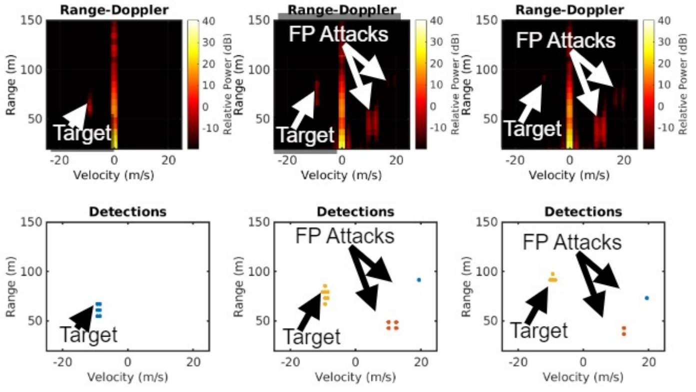
Vulnerability Analysis of mmWave Radars

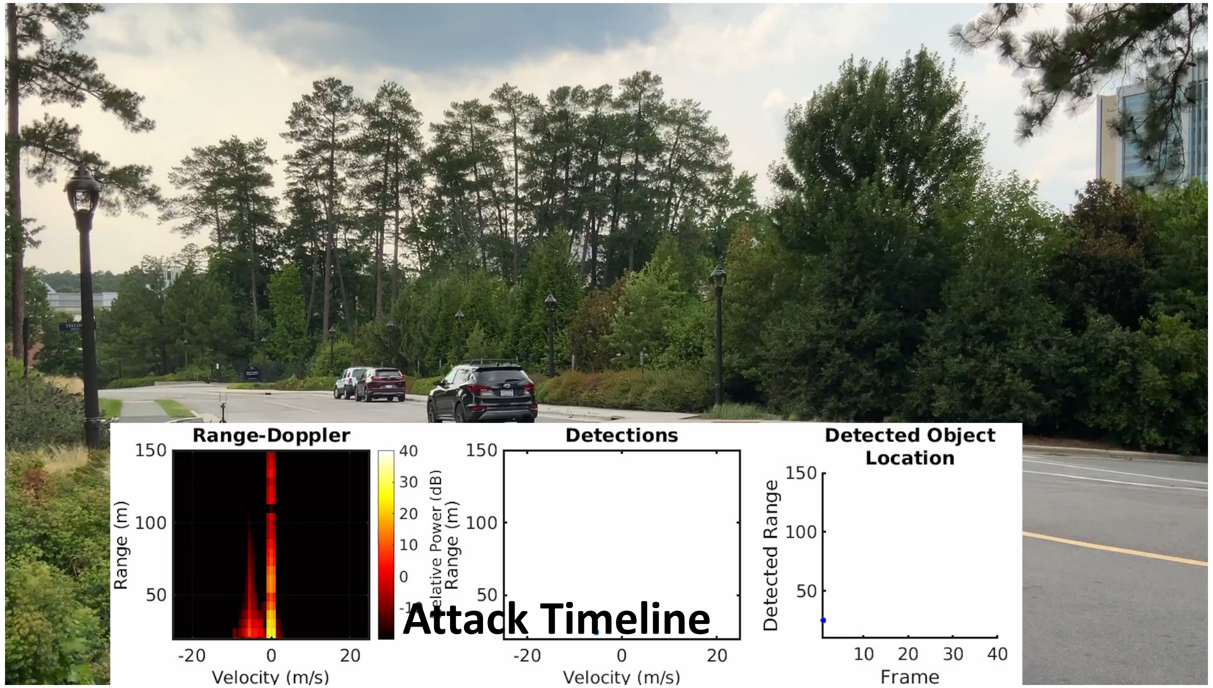
MadRadar: A Black-Box Physical Layer Attacks (NDSS'24)

Duke PRATT SCHOOL of ENGINEERING

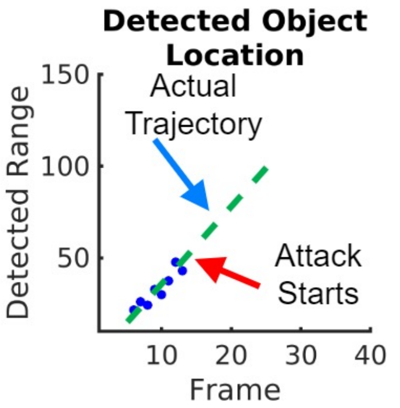
False Positive Attacks

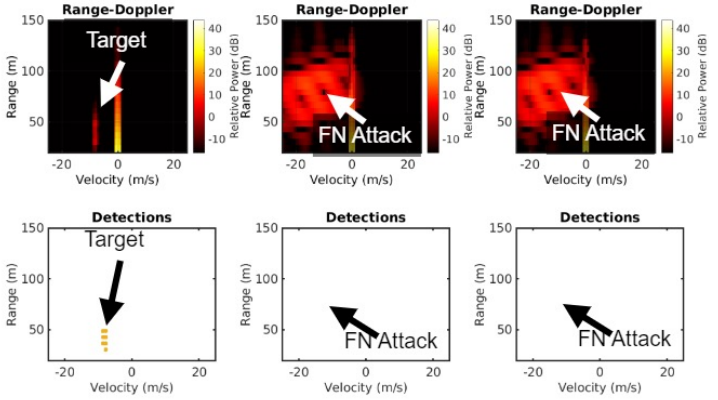


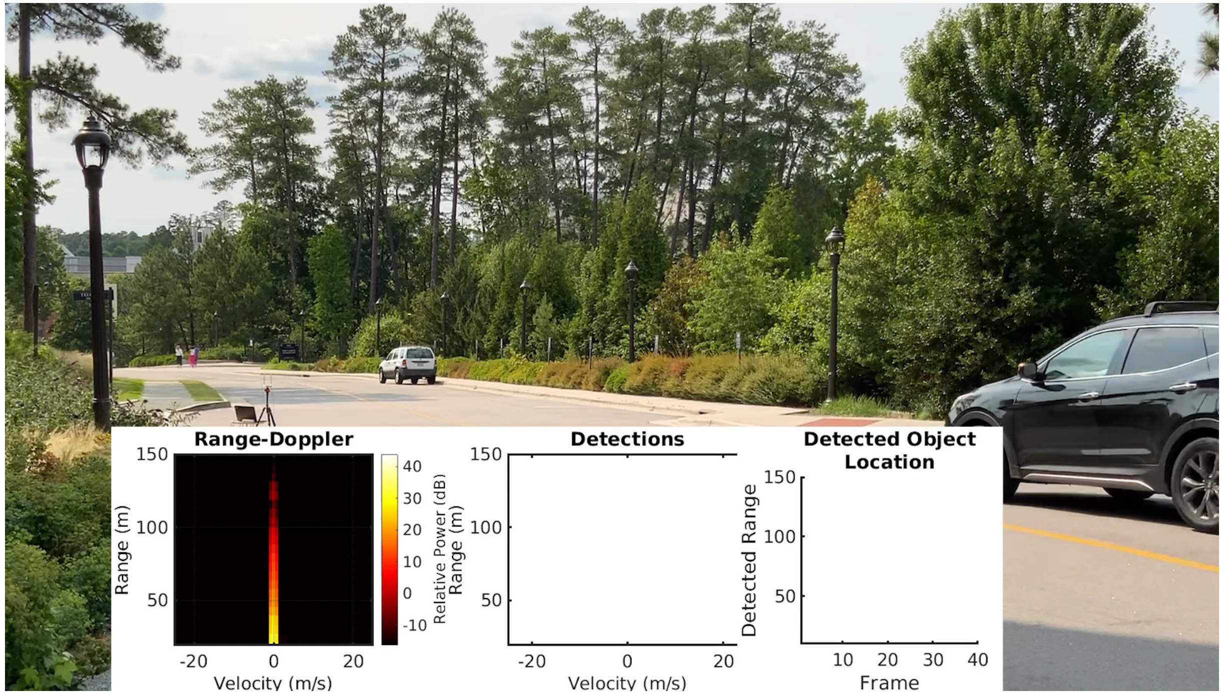




False Negative Attacks

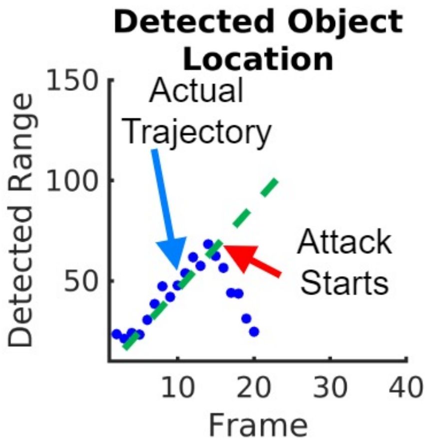




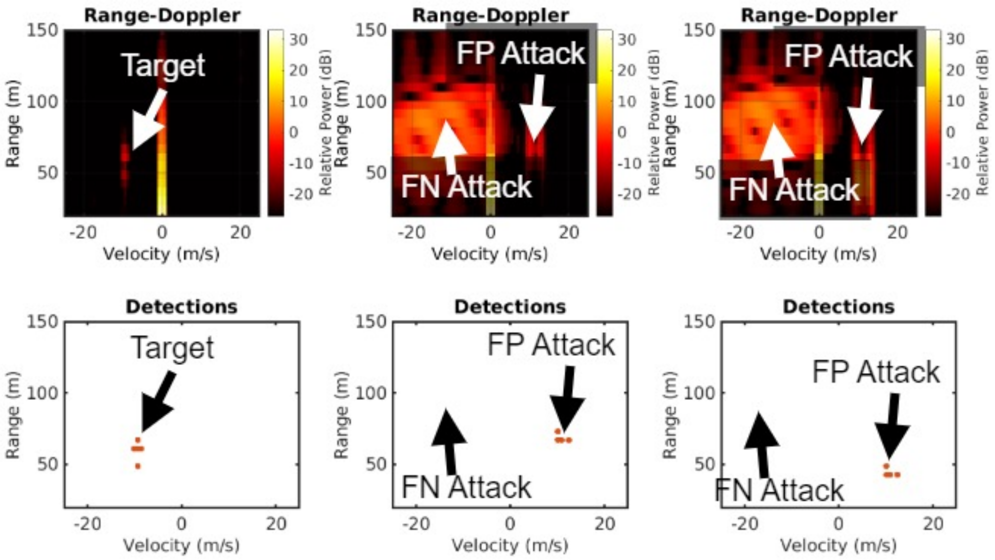


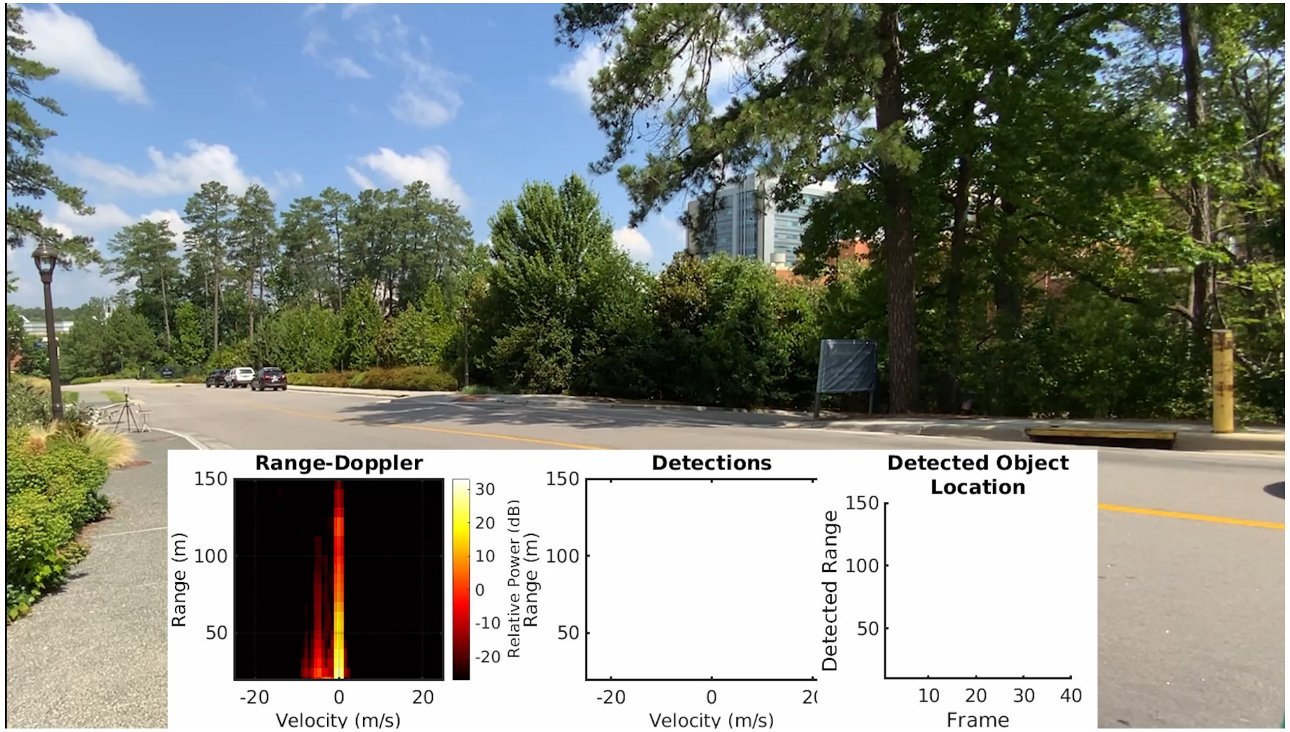
Vulnerability Analysis of mmWave Radars
MadRadar: A Black-Box Physical Layer Attacks (NDSS'24)

Translation Attacks



Attack Timeline



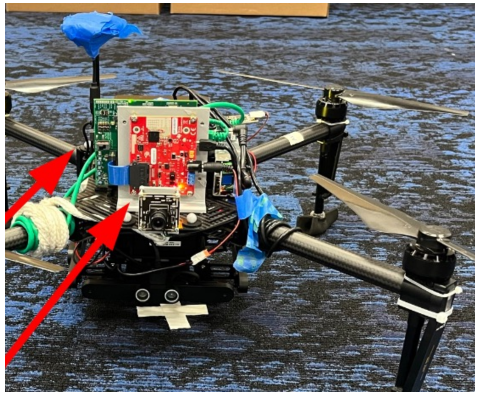


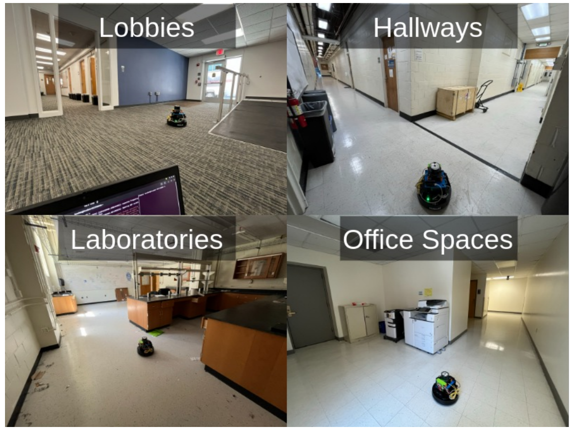
https://sites.google.com/view/madradar-public/home

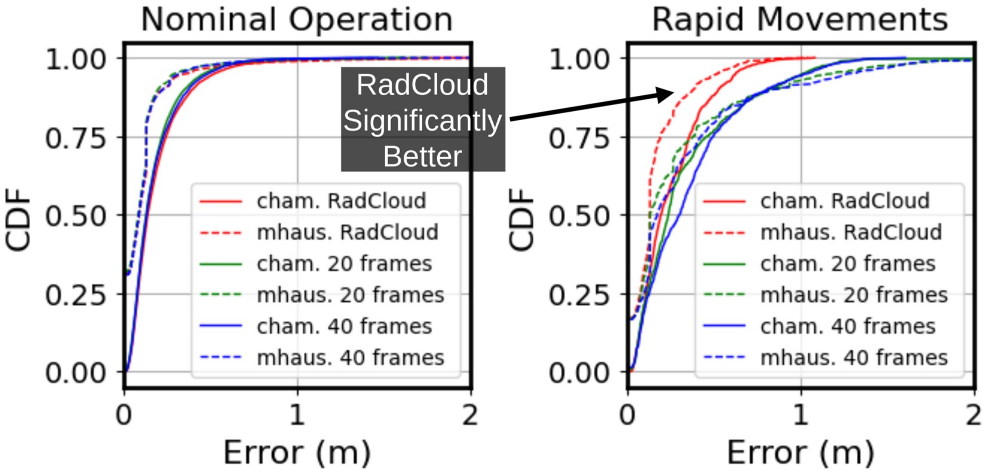
mmWave-based Autonomy (ICRA’24*)

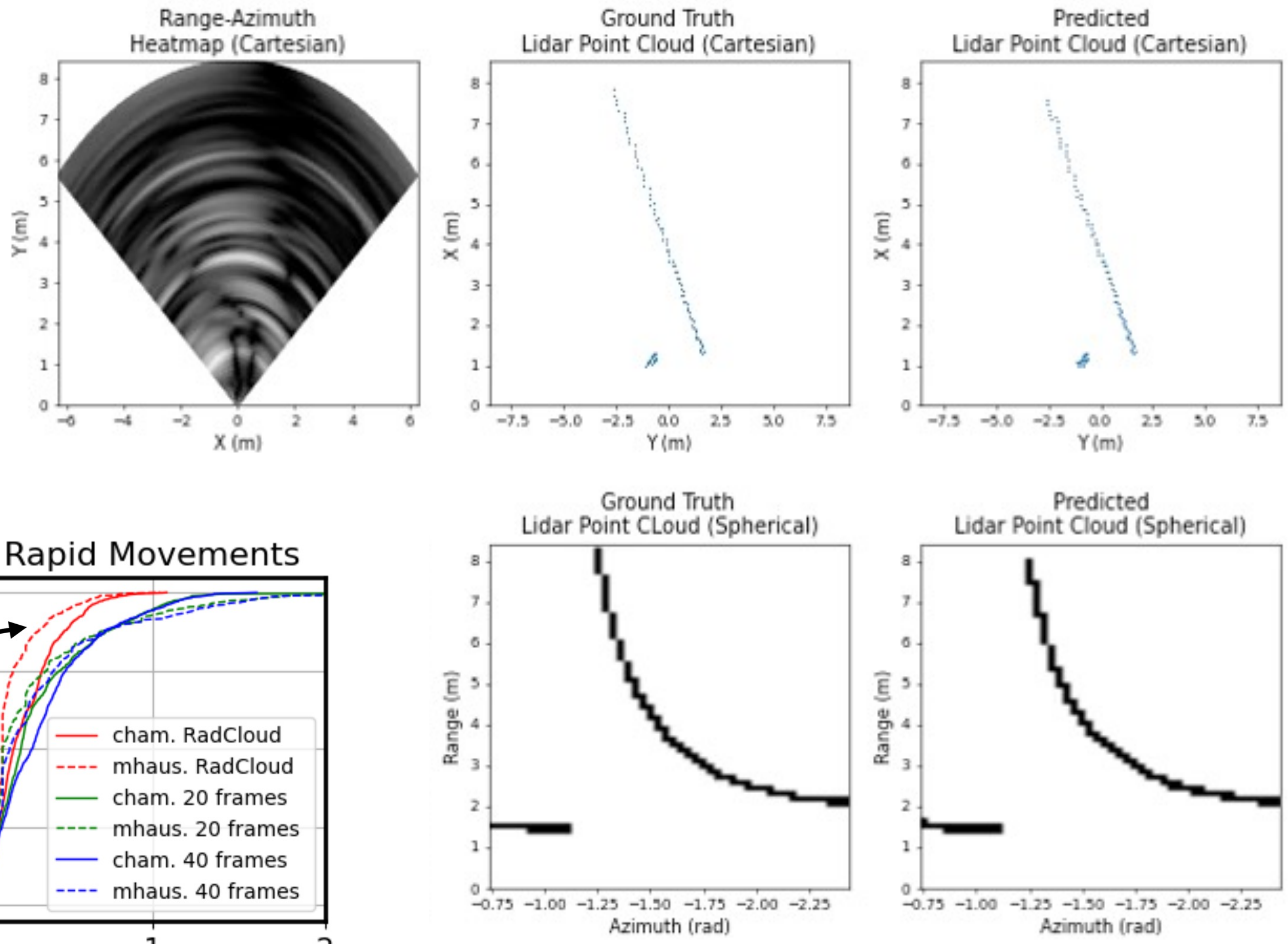
Goal: ***Low-cost*** (~$100), ***low-weight*** solution for *adversarially robust* situational awareness and autonomy on ***computationally constrained*** devices

- ***Real-time*** *high-accuracy* point clouds on a NUC-powered drone using mmWave sensing

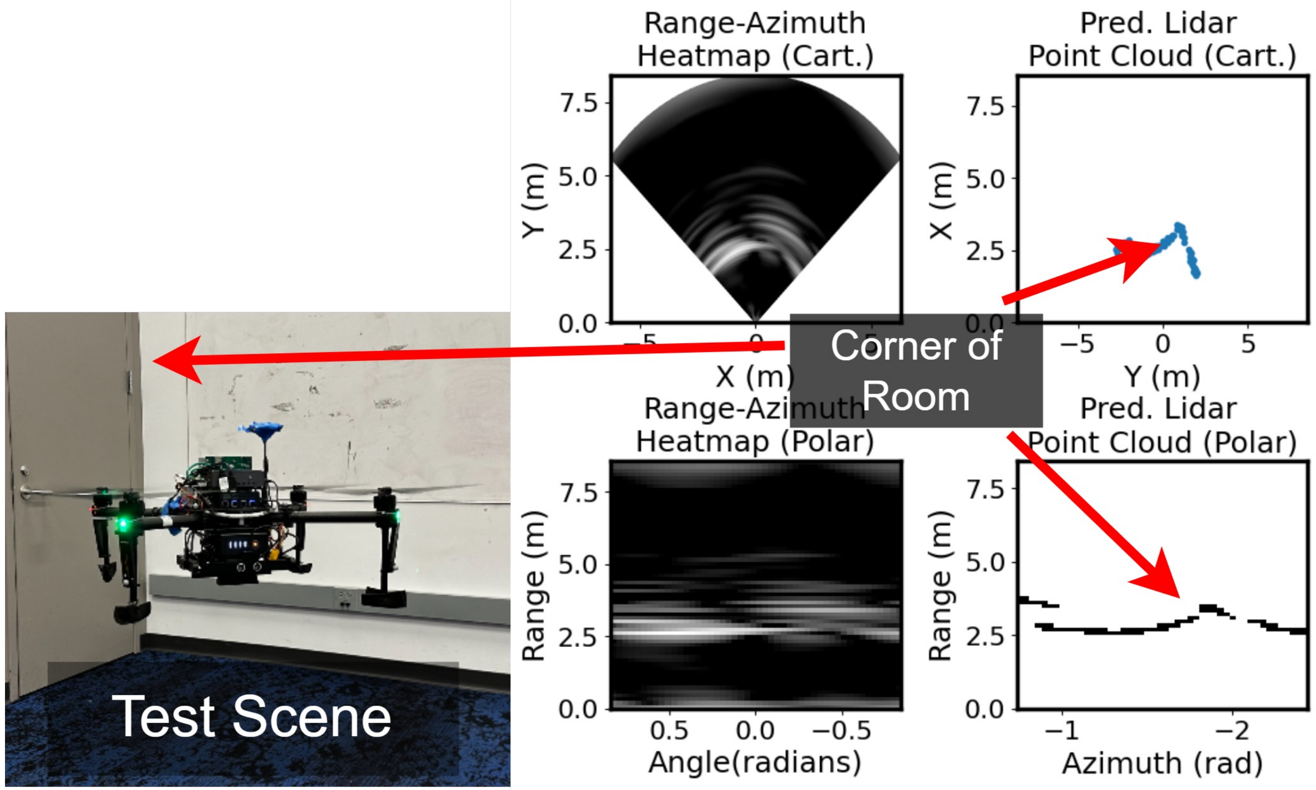




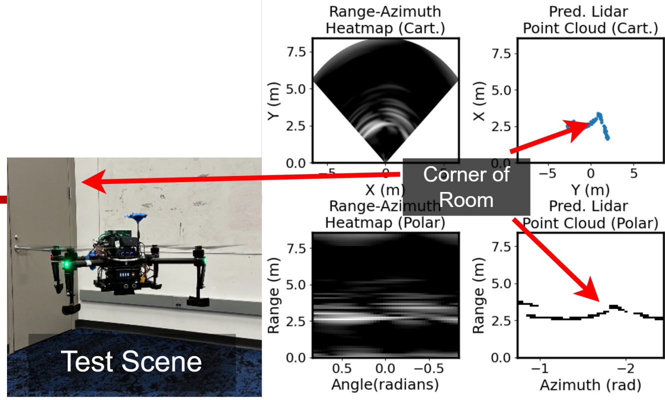




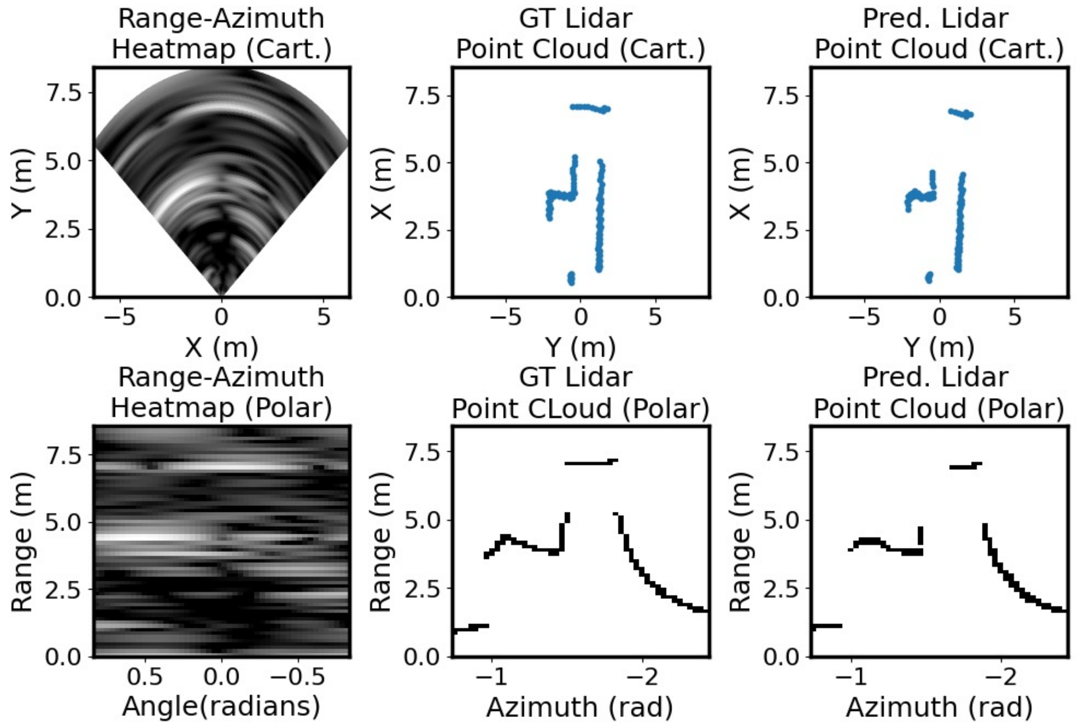
mmWave-based Autonomy (ICRA’24*)



mmWave-based Autonomy (ICRA'24*)

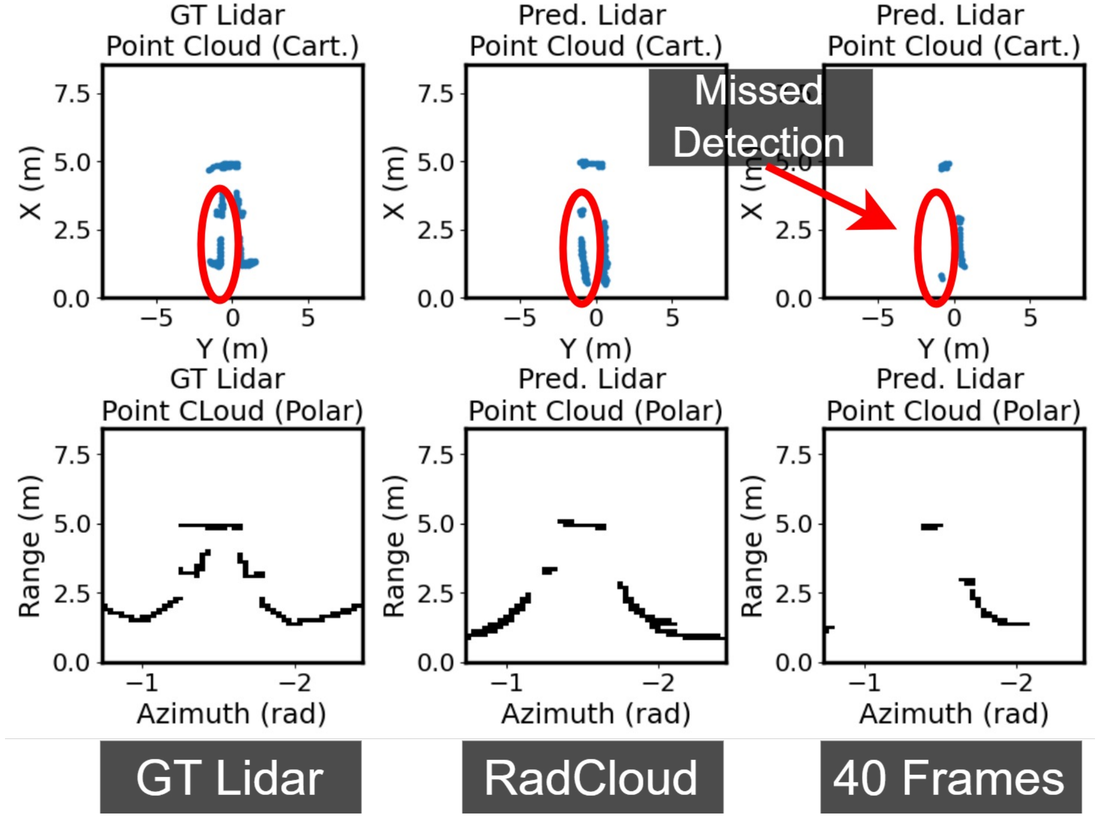


New Environments



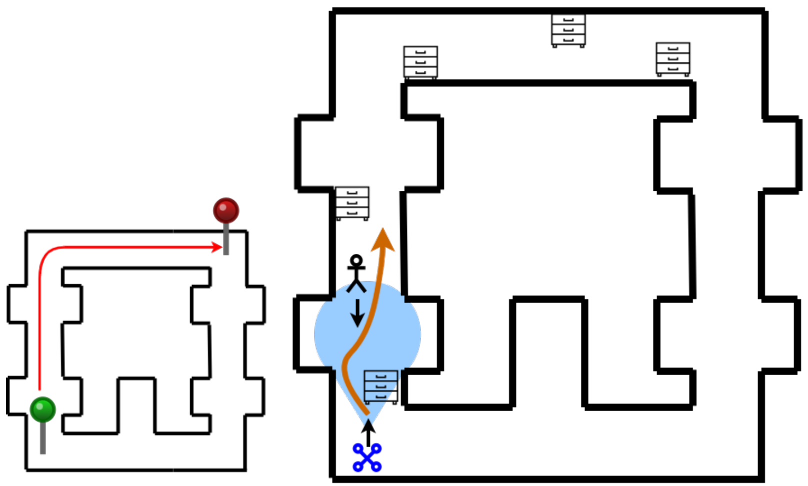
Input radar data, ground truth point cloud, and predicted point cloud

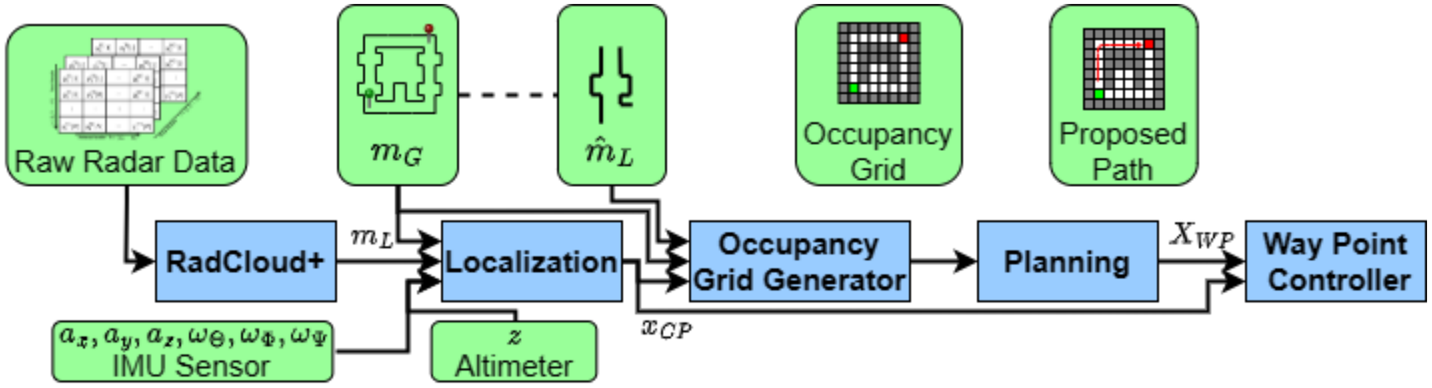
Aggressive Maneuvers



RadNav

- *Goal: N*avigate through an environment using only radar and traditional odometry sensors





Thank you