USENIX Symposium on Vehicle Security and Privacy 2025

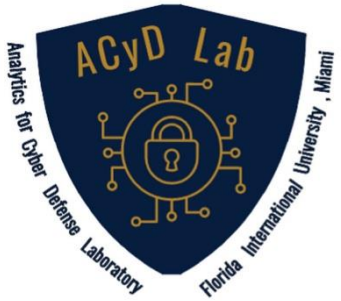
SLEIGHT: Real-Time Closed-Loop Attack Analytics Framework for Holistic Intrusion Detection Systems in Unmanned Aerial Vehicles

Muneeba Asif[1], Maryna Veksler[2], Devrim Aras[1], Kemal Akkaya[2], Mohammad Ashiqur Rahman[1]

[1]Florida International University, [2]Virginia Commonwealth University





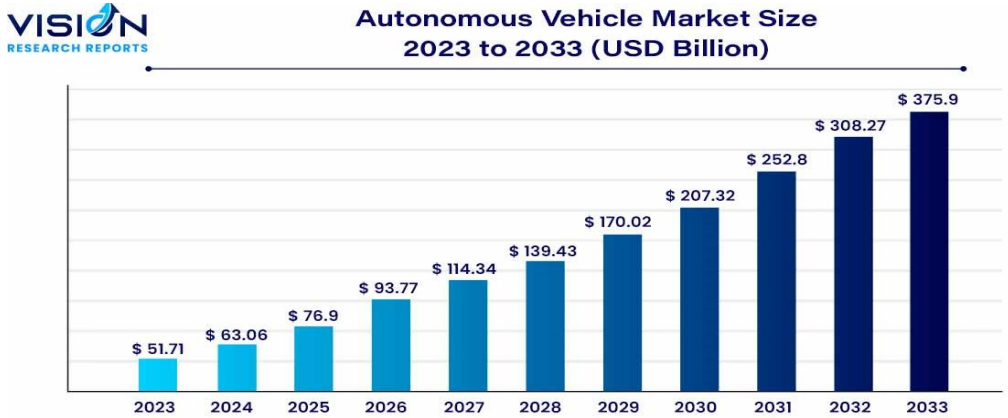






Introduction

- The unmanned aerial vehicle (UAV) industry is projected to exceed $97.65 billion by 2031[1,2].
- UAVs are integral in varied applications such as
  - surveillance for smart grids
  - crowd analysis
  - disaster and urban monitoring
  - agriculture
  - remote sensing
  - logistics and
  - defense sectors
- To proficiently handle these mission-critical tasks, UAVs must ensure end-to-end security and resiliency without sacrificing safety.



Disaster Management



Power Grid Inspections





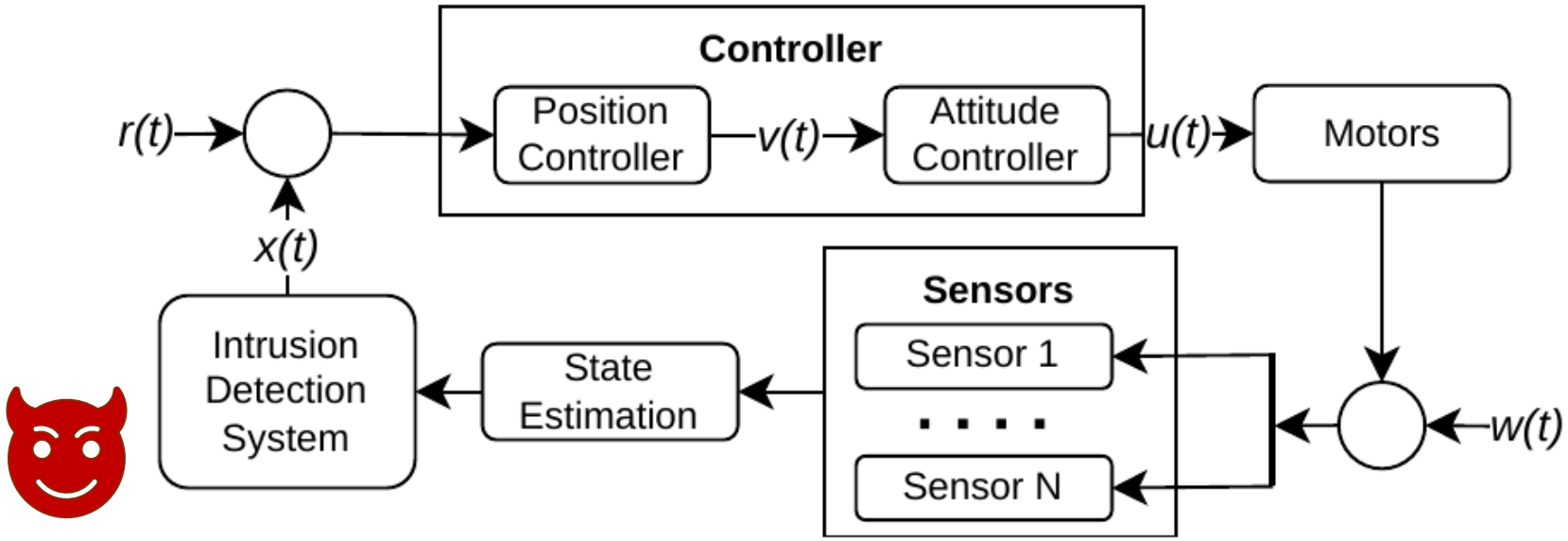
Precision Agriculture



Defense and Military

[1] https://droneii.com/drone-market-analysis-2022-2030, Sep 2022.
[2] https://www.visionresearchreports.com/autonomous-vehicle-market/41070

Background: Closed-Loop UAV Operation



Sensing → estimation → control → actuation, every control cycle.

Robustness against an adaptive, real-time attacker inside this loop is uncharacterized.

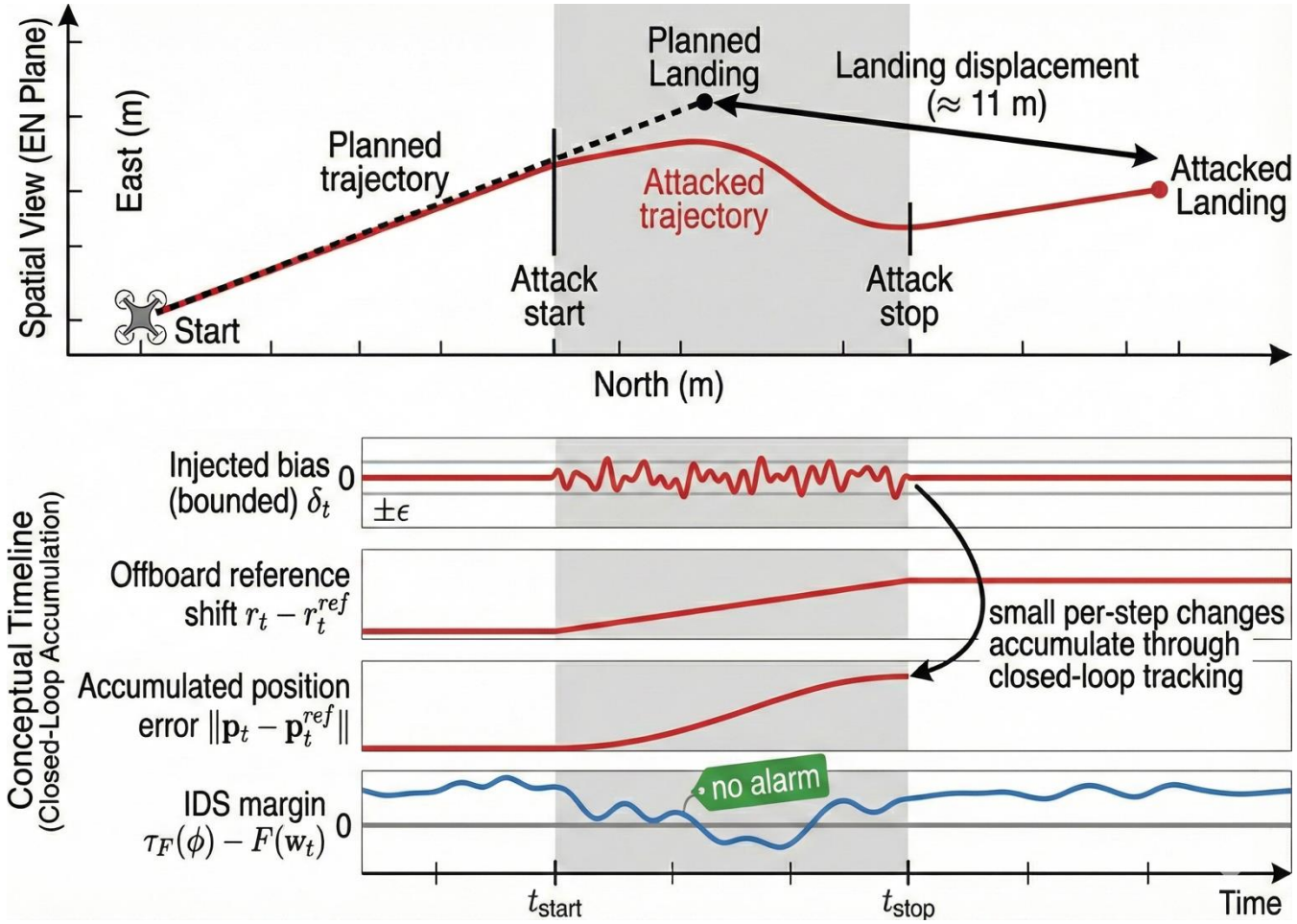
State-of-the-Art

	Closed-loop in flight	Real-time synthesis	Physically plausible	Holistic IDS evaluated
Sensor spoofing (GPS, acoustic, IEMI)	○	○	○	○
Adversarial ML on IDS (FGSM, PGD, GAN)	○	○	○	◐
CPS stealth–impact / falsification	○	○	◐	○
Online RL evasion of log detectors	◐	●	○	○
SLEIGHT	●	●	●	●

● yes ◐ partial ○ no

Offline and open-loop evaluation cannot accurately answer what an in-flight attacker can actually do.

Motivating Example

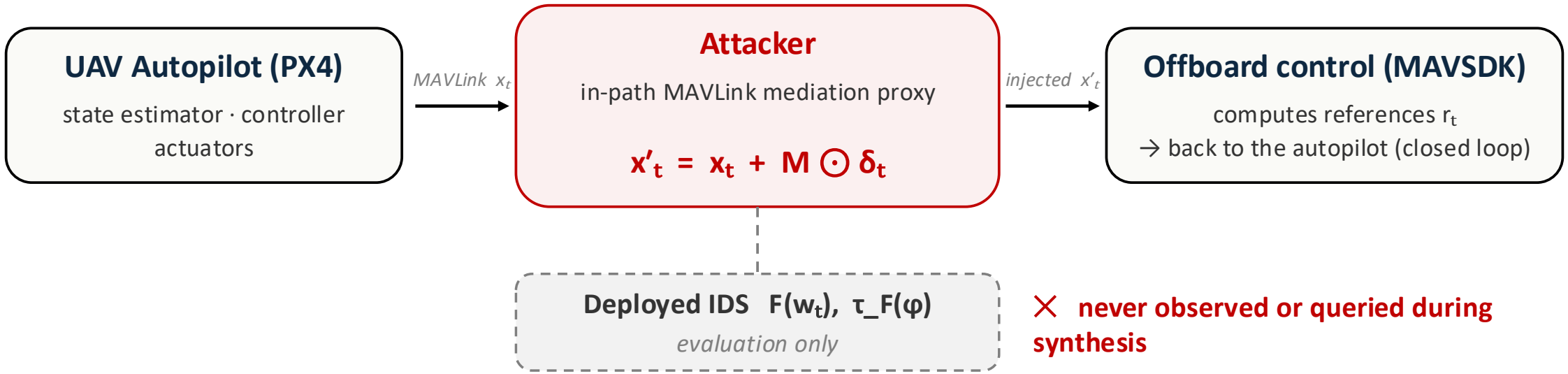


- Each perturbation stays inside the detector's tolerance.
- Closed-loop tracking accumulates the bias.
- **≈ 11 m landing displacement.**

How much mission deviation can an attacker induce while staying nominal-looking, over an entire flight?

*IDS = Intrusion Detection System

Threat Model



Attackable channels (mask M)

GPS position · LOCAL_POSITION_NED · ATTITUDE · barometric altitude

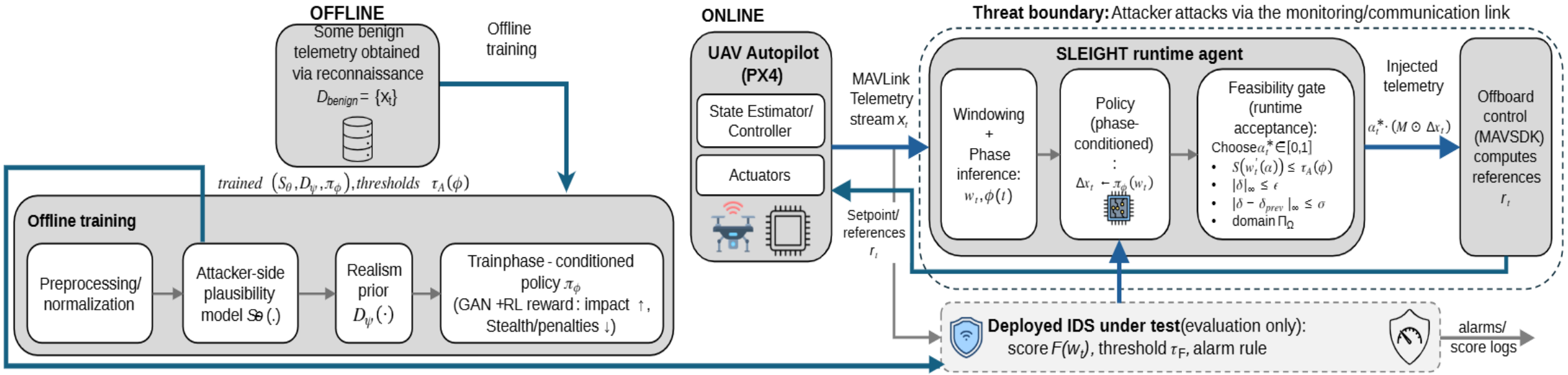
single-channel, or coordinated to preserve estimator consistency

Never modified

Firmware / safety logic· controller variables · estimator status

firmware · mission plan · safety logic · keys · no RF denial

SLEIGHT Overview



Offline - from reconnaissance only, we train

- Plausibility scorer Sθ (IDS* residual)
- Discriminator Dψ (cGAN* realism critic)
- Conditional generator / policy πφ

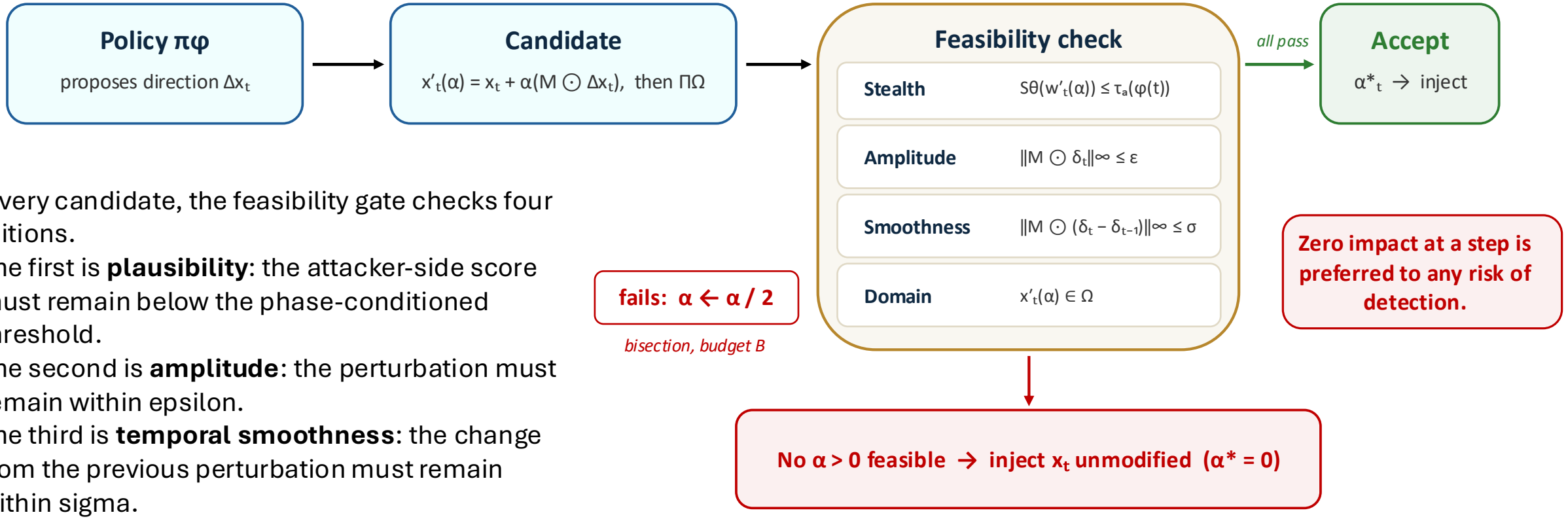
Online - for every control step

- Policy proposes a direction Δx_t
- Feasibility gate picks the scale α^*_t
- Inject, then MAVLink forwards it on

The deployed IDS is never queried, F(·) is used for evaluation only.

*IDS = Intrusion Detection System
*cGAN = Conditional GAN

Runtime Feasibility Enforcement



For every candidate, the feasibility gate checks four conditions.

- The first is **plausibility**: the attacker-side score must remain below the phase-conditioned threshold.
- The second is **amplitude**: the perturbation must remain within epsilon.
- The third is **temporal smoothness**: the change from the previous perturbation must remain within sigma.
- And the fourth is **domain validity**, so the resulting telemetry remains valid for its physical and message-level range.

SLEIGHT uses bisection to find the **largest alpha satisfying all four constraints**.
When the full proposal is too aggressive, alpha decreases. When no positive alpha satisfies the constraints, the original telemetry is forwarded for that step.

Policy Training: cGAN – RL Model

Conditional Generative Adversarial Network (cGAN)

Generator πφ

conditioned on $w_t \cdot \varphi(t) \cdot \Delta x_{t-1}$
plus noise $z \sim N(0, I)$

Δx_t

Discriminator Dψ

benign windows vs. bounded perturbed windows

→ realism score $R_real = \log D\psi(w'_t)$ feeds the reward

Reinforcement Learning Reward

Deviation Term:

$$\Delta(\mathbf{x}_t^{\text{adv}}, \mathbf{x}_t) = \|\mathbf{x}_t^{\text{adv}} - \mathbf{x}_t\|$$

Penalty for Unrealistic Changes:

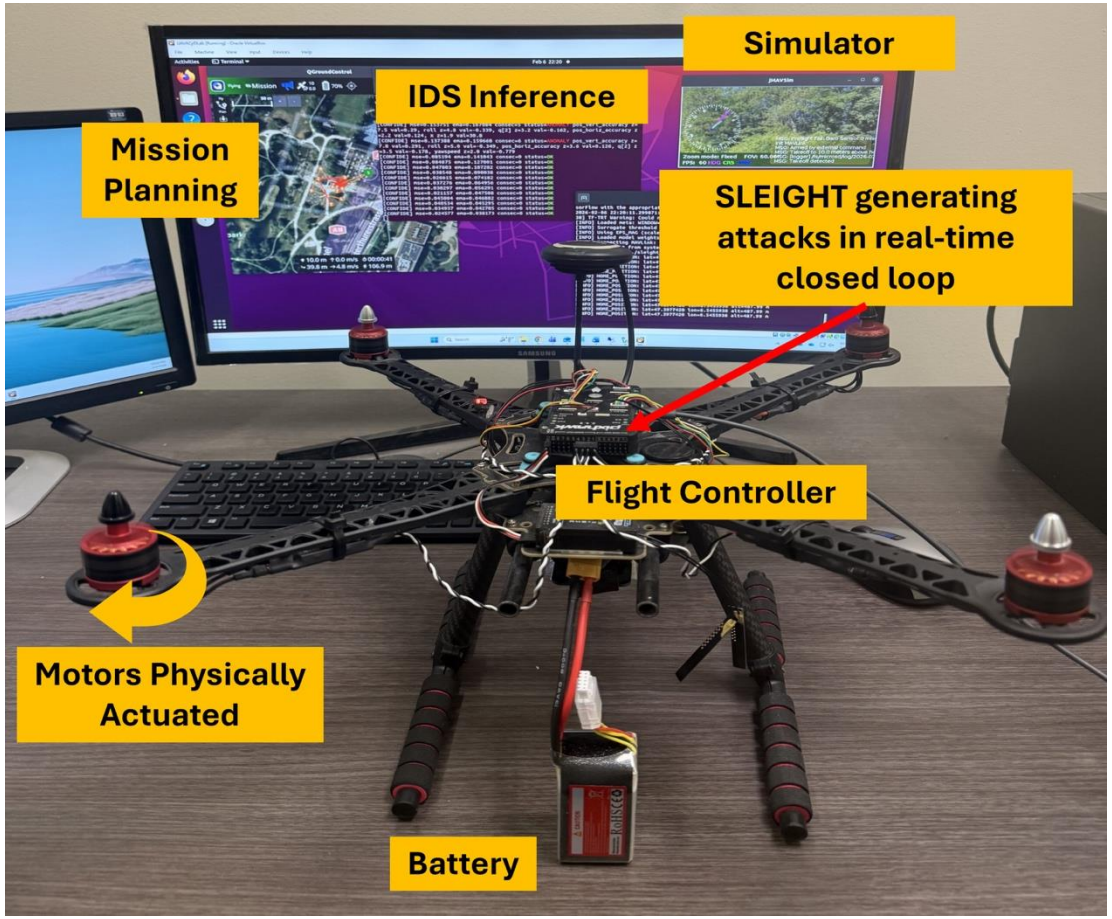
$$\Psi(\mathbf{x}_t^{\text{adv}}) = \max\{0, \|\mathbf{x}_t^{\text{adv}} - \mathbf{x}_t\| - \delta_{\max}\}$$

Total Reward:

$$\mathcal{R}(\mathbf{x}_t^{\text{adv}}, \mathbf{x}_t) = \begin{cases} 1 + \alpha \cdot \Delta(\mathbf{x}_t^{\text{adv}}, \mathbf{x}_t) - \beta \cdot \Psi(\mathbf{x}_t^{\text{adv}}), & \text{if } \varepsilon(\mathbf{x}_t^{\text{adv}}) \le \theta, \\ 0, & \text{otherwise} \end{cases}$$

- $\alpha \cdot \Delta(\mathbf{x}_t^{\text{adv}}, \mathbf{x}_t)$ — Maximum Attack
- $\beta \cdot \Psi(\mathbf{x}_t^{\text{adv}})$ — Within realistic limits
- $\varepsilon(\mathbf{x}_t^{\text{adv}}) \le \theta$ — remaining under the threshold

Implementation and Evaluation



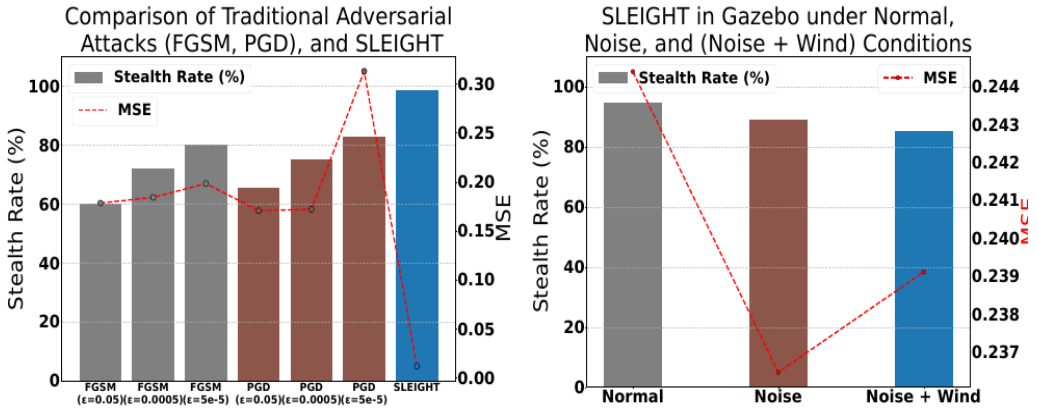
- Pixhawk runs PX4, streams MAVLink over telemetry link. SLEIGHT proxy mediates the MAVLink autopilot-to-offboard control stream on companion host.
- Message IDs, encodings, and stream rates preserved, i.e., 10 Hz navigation, 5 Hz GPS, unchanged during injection.
- **Metrics:** evasion (no-alarm rate), attack success (impact), perturbation magnitude, and mission impact measured in closed loop.

Mission and Conditions

RQ1 Does SLEIGHT generalize beyond a single mission profile?

Flight plan	Evasion %	Attack success %
Base (9 WPs)	100	85.9
Zig Zag (15)	97.9	86.5
Fig-8 (20)	96.9	85.2
Spiral Ascent (25)	100	85.1

- **Regardless of mission duration, evasion > 96% at every duration.**
- Altitude jumps every other waypoint: stealth still ≥ 98.5%.
- Wind (8 m/s) + GPS noise: stealth 80–85%, and the policy adapts.



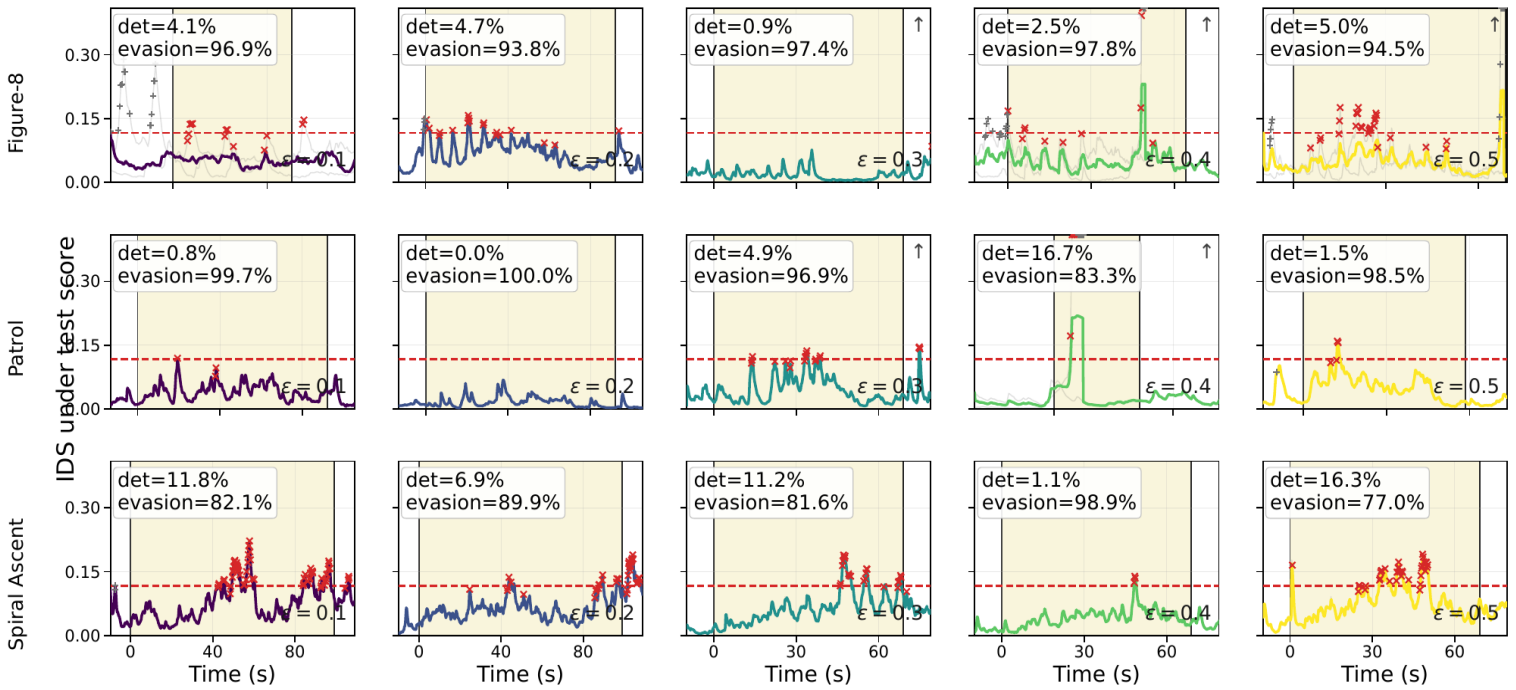
SLEIGHT vs. FGSM / PGD, and under wind + noise.

Why not just FGSM / PGD?

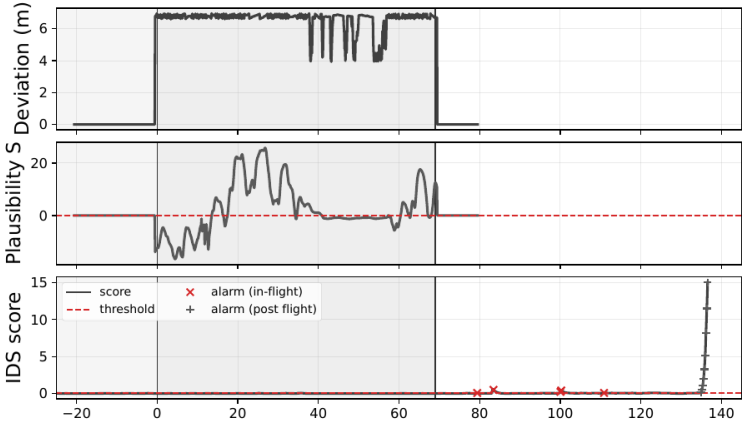
- Small gradient steps cause abrupt velocity and heading changes.
- They propagate through the loop, hence, destabilizing flight, sometimes crashing.
- **Loss of stability undermines stealth in closed-loop operation.**

HITL Validation

RQ2 Does the closed-loop behavior transfer to real flight-controller firmware?



Deployed IDS score over time: three missions × five budgets ε.



One HITL run: deviation, plausibility S, IDS score.

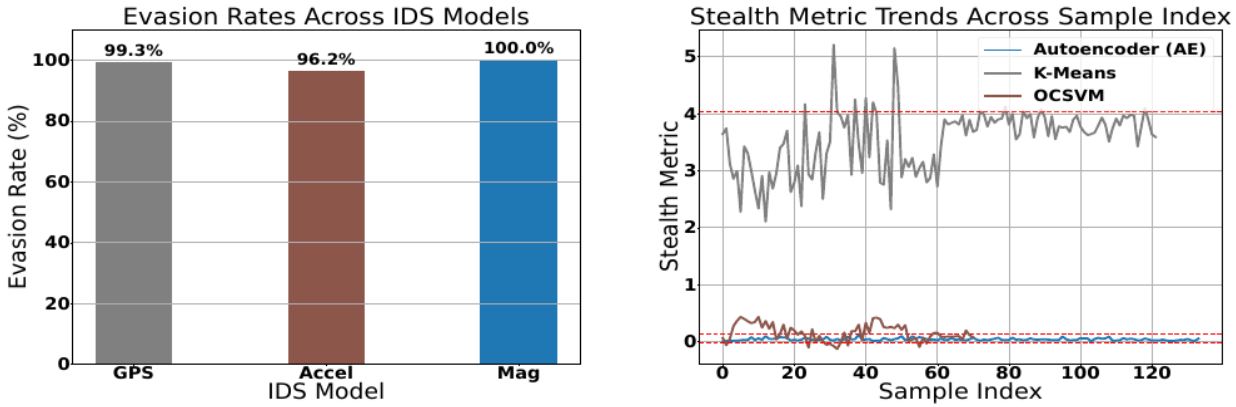
- Evasion 77–100%; most configurations above 90%.
- **ε ∈ {0.2, 0.3}: ≥ 93.8% on all three missions.**
- ε = 0.5 costs stealth on long missions (16.3% detection).

The stealth–impact tradeoff persists under real firmware timing, sensor noise, and estimator dynamics.

IDS Architectures

RQ3 Does stealth and impact generalize beyond a single detector design?

IDS Architecture	Evasion %
Autoencoder	≈ 99 - 100
K-Means	90.2
OCSVM	87.3
ConFIDe (holistic, control-fused)	>95
PID-Piper (control-aware, CUSUM)	>90

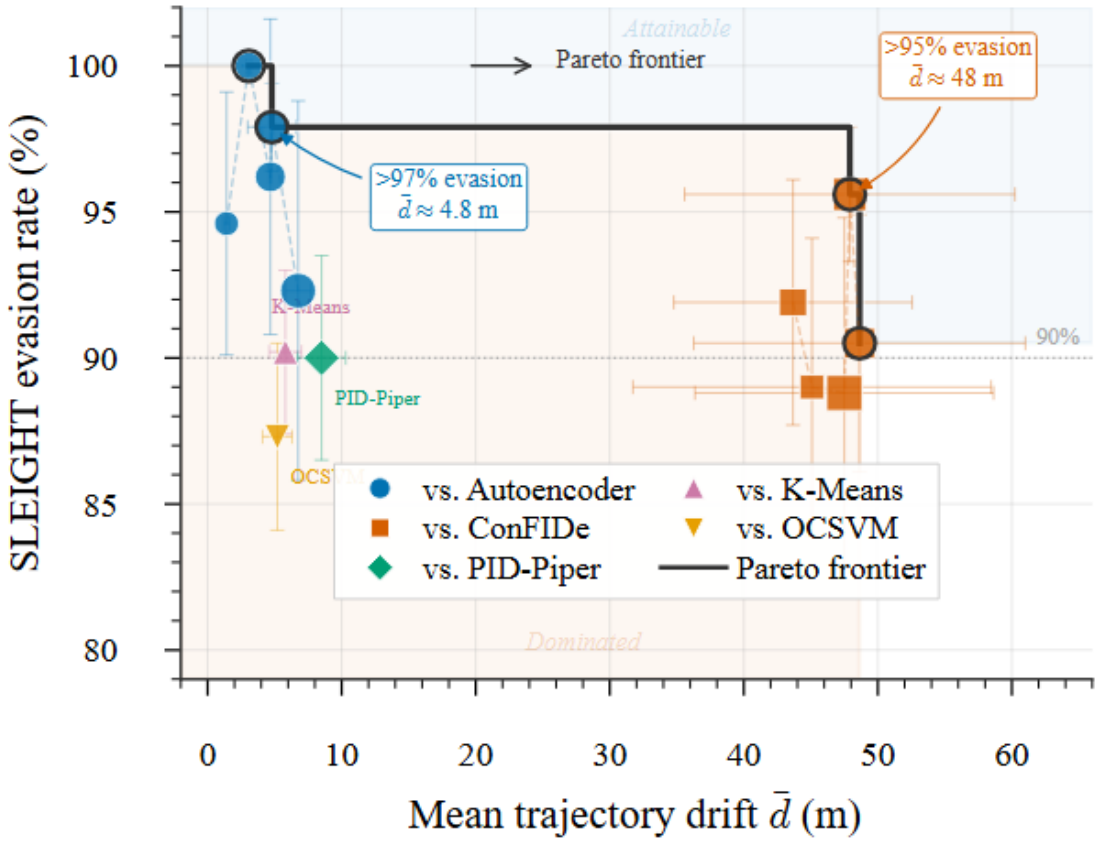


Evasion and stealth trends across ML-based IDSs.

- Reconstruction-based detectors are the most exposed, the realism prior targets exactly the manifold they model.
- Boundary and residual detectors shrink the feasible space, but do not close it.

Stealth-Impact Frontier

RQ4 What mission impact is achievable at each stealth level?



> 97% evasion

đ ≈ 4.8 m

> 95% evasion

đ ≈ 48 m

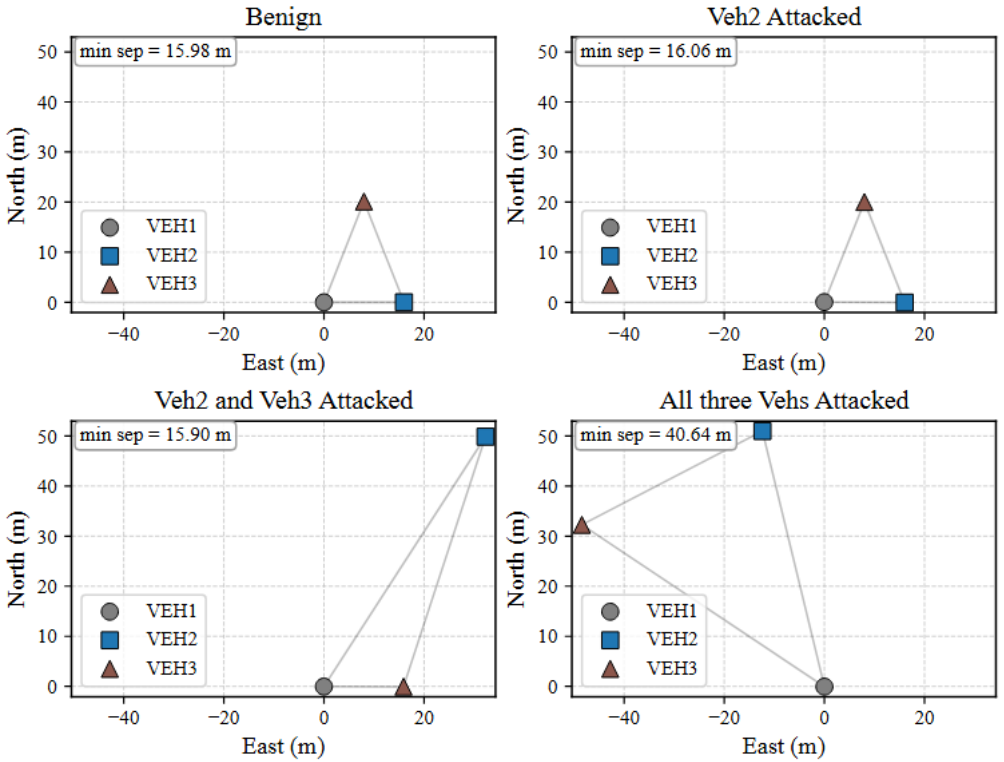
≈ 2%-point stealth reduction

↓

≈ 10× mean drift

Swarm Analysis

RQ5 Does the attack scale to coordinated multi-UAV operations?



Swarm positions (local EN); dashed line marks sep_min = 10 m.

Landing displacement from target (m)

Attacked	Veh 1	Veh 2	Veh 3
Benign	0.04	0.09	0.08
Veh-2	0.04	**52.5**	21.7
Veh-2 & 3	0.05	**58.3**	**57.7**
All three	**60.1**	**88.3**	**61.0**

- Per-vehicle stealth preserved, IDS scores stay well below τ_eff.
- Impact is selective: untouched vehicles keep tracking nominally.
- **Formation stays safe, d_min ≥ 10 m throughout.**

Conclusion

- Bounded, physically plausible per-step perturbations accumulate into mission-level deviation under closed-loop feedback.
- Offline and open-loop evaluation overestimates IDS robustness; behavior differs under closed-loop execution.
- All four IDS families were evaded, including holistic controller- and actuator-aware designs: ConFIDe > 95%, PID-Piper > 90%.
- Frontier: > 97% evasion at $\bar{d} \approx 4.8$ m; > 95% evasion at $\bar{d} \approx 48$ m. Runtime 1.9×10^{-4} s per step on CPU, within the 0.1 s budget at 10 Hz

Countermeasures

- Adaptive or randomized thresholds: Our threshold-sensitivity analysis shows degraded stealth and irregular drift when the boundary moves.
- Cross-sensor and physics checks, i.e., EKF innovation and residual consistency across GPS, IMU, magnetometer, estimator and control signals.
- Online updates and ensembles avoids a single fixed decision boundary for the attacker to track.
- Closed-loop adversarial calibration, i.e., use SLEIGHT traces to calibrate or adversarially train against these failure modes.



Thank you for Listening!

This research was supported in part by the Department of Energy (DOE) under awards DE-CR0000024 and DE-CR0000046.

Questions?

