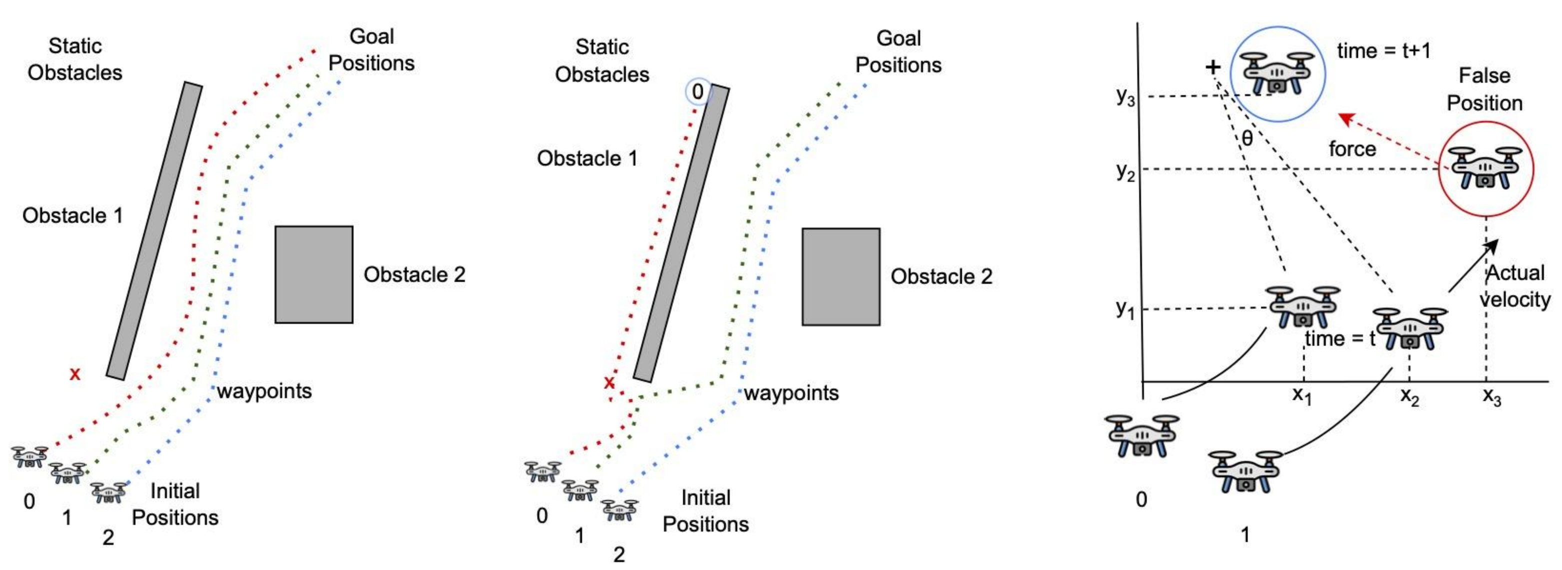
Raven: Automated Discovery of Semantic Attacks in Multi-Agent Navigation Systems

Doguhan Yeke†, Kartik A. Pant†, Muslum Ozgur Ozmen‡, Hyungsub Kim§, James M. Goppert†, Inseok Hwang†, Antonio Bianchi†, and Z. Berkay Celik†

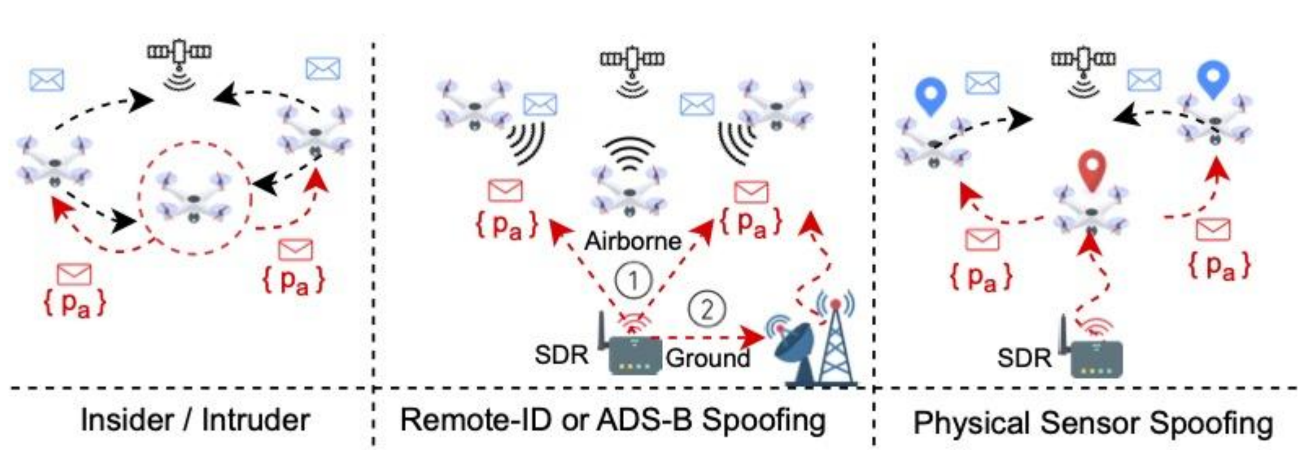
†Purdue University, ‡Arizona State University, §Indiana University Bloomington

Introduction

- Autonomous multi-robots (AMRs) rely on collision avoidance algorithms for surveillance, logistics, and security operations.
- However, these systems are vulnerable to False Data Injection Attacks (FDIAs).
- Existing methods fail to consider complex multi-robot dynamics and the full spectrum of attacks.



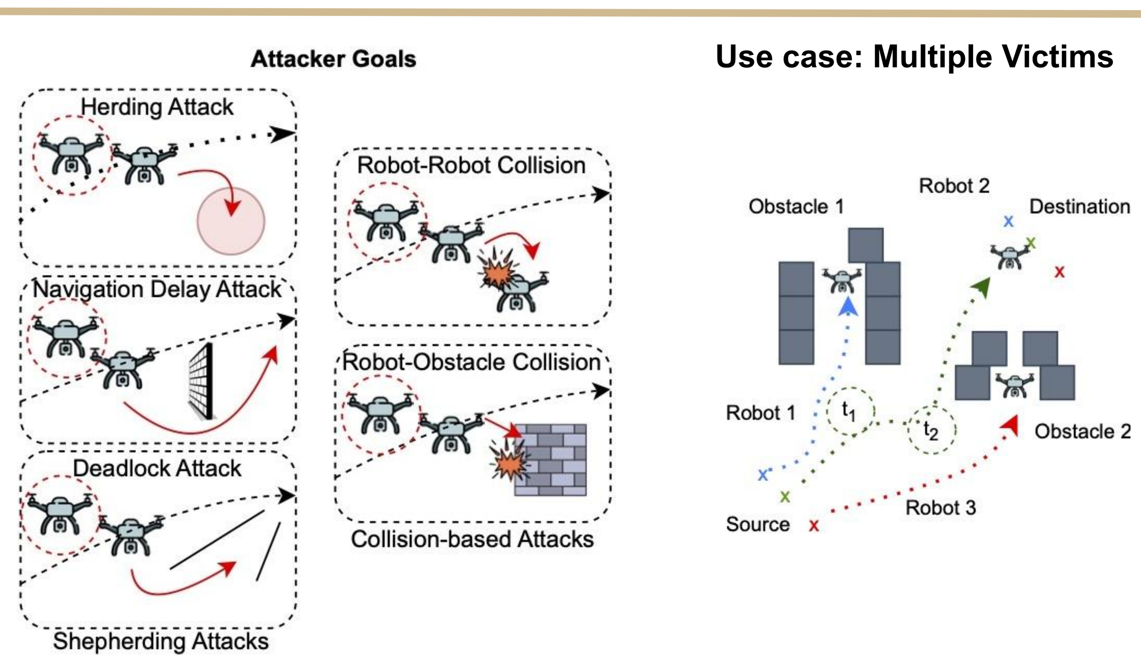
Threat Model



- Insider/Intruder: A malicious robot within the swarm injects false position data into the network.
- Remote ID/ADS-B Spoofing: The attacker exploits unauthenticated and unencrypted broadcast protocols to transmit fake robot locations.
- Sensor Spoofing: The attacker spoofs GNSS signals, causing the target robot to report an incorrect position.

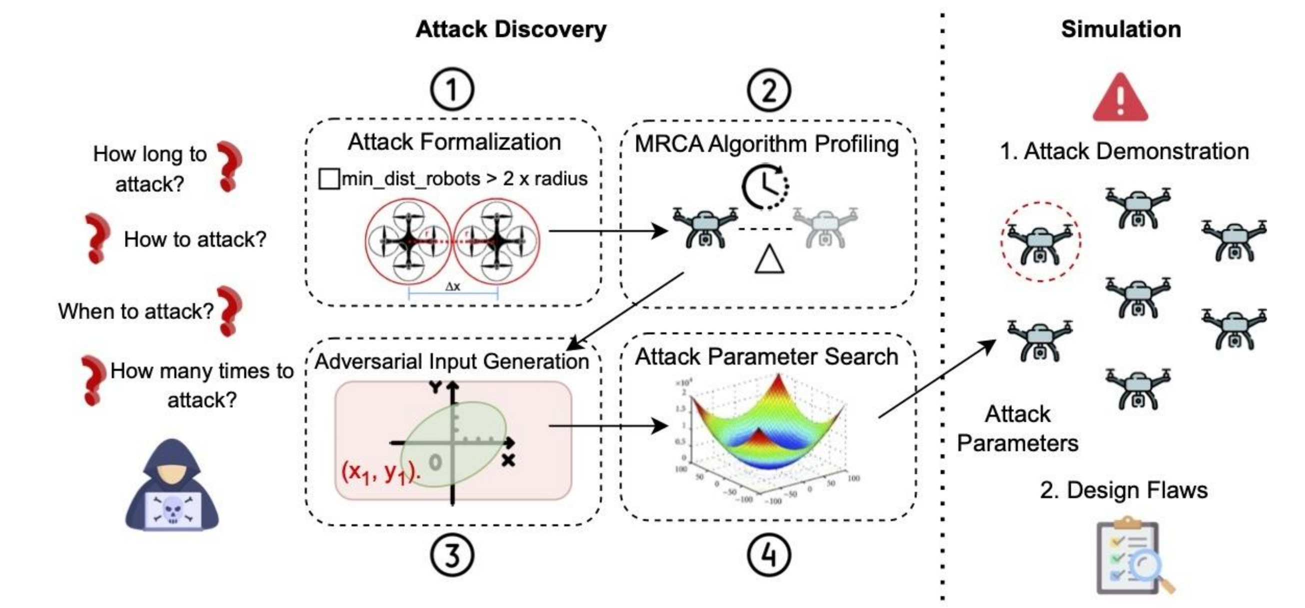
Attacker Goals

- Herding: Forcing a victim into attacker-defined area.
- Deadlock: Immobilizing robots for certain time.
- Navigation Delay: Forcing a victim to take a longer, suboptimal route.
- Robot-Robot Collision: Inducing collisions between robots.
- Robot-Obstacle Collision: Inducing collisions with obstacles.



RAVEN Overview

- Uses Signal Temporal Logic (STL) for formal attack specification.
- MRCA algorithm profiling.
- Employs stochastic optimization for finding stealthy attack parameters.
- Minimize detection by maintaining spatio-temporal consistency and sensor noise ranges.



Evaluation Results

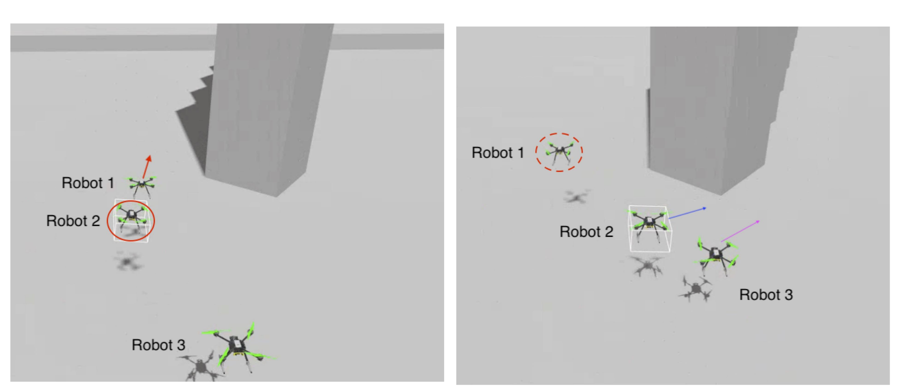
Attack Goal	Benign Case	Attack Discovery	Min # Injections	Attack Plan Time	Root Cause†
Experiments on ORCA					
Robot-Robot Collision	0/10 (0%)	10/10 (100%)	1	2.38 s / 2.6 s / 2.94 s	HR-ICM-PTT-FC
Robot-Obstacle Collision	0/10 (0%)	10/10 (100%)	1	2.5 s / 4.2 s / 4.6 s	HR-ICM-PTT-FC
Herding	0/10 (0%)	10/10 (100%)	1	1.97 s / 2.26 s / 2.53 s	HR-ICM-PTT
Deadlock	0/10 (0%)	10/10 (100%)	1	1.22 s / 2.2 s / 2.44 s	HR-ICM-PTT
Navigation Delay	0/10 (0%)	10/10 (100%)	1	1.01 s / 3.35 s / 5.63 s	HR-ICM-PTT
Experiments on GLAS					
Robot-Robot Collision	0/10 (0%)	10/10 (100%)	1	7:39 s / 7:58 s / 8:54 s	ICM-PTT-LA
Robot-Obstacle Collision	0/10 (0%)	10/10 (100%)	1	8:4 s / 10:2 s / 14:8 s	ICM-PTT-LA
Herding	0/10 (0%)	9/10 (90%)	3	2:35 s / 2:4 s / 2:42 s	ICM-PTT-LA
Deadlock	0/10 (0%)	10/10 (100%)	3	1:54 s / 2:44 s / 2:52 s	ICM-PTT-LA
Navigation Delay	0/10 (0%)	10/10 (100%)	3	2:15 / 2:22 s / 2:36 s	ICM-PTT-LA

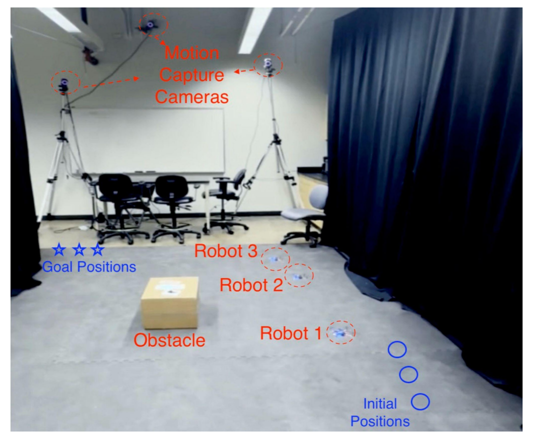
† HR: High Reactivity, ICM: Imperfect Communication and Measurements, PTT: Planning vs. Time Tradeoff, LA: Learning-based Algorithms, FC: Feasibility of Collisions.

- Successfully identified semantic attacks against ORCA and GLAS algorithms.
- Demonstrated stealthiness by evading anomaly detection mechanisms.
- Conducted experiments using high-fidelity simulator and Crazyflie drones demonstrating practicality of attacks.

Root Causes:

- High Reactivity
- Imperfect Communication
- Planning vs. Time Tradeoff
- Learning-based Flaws
- Feasibility of Collisions





Conclusion

- Discovered new semantic attack scenarios in multi-robot navigation.
- Introduced Raven framework to systematically uncover vulnerabilities.
- Identified key design flaws in widely adopted MRCA algorithms.
- Suggested robust countermeasures for enhancing system resilience.

Real-world End-to-end Demonstration

References

[1] Doguhan Yeke, Kartik A. Pant, Muslum Ozgur Ozmen, Hyungsub Kim, James M. Goppert, Inseok Hwang, Antonio Bianchi, and Z. Berkay Celik. Automated Discovery of Semantic Attacks in Multi-Robot Navigation Systems. Usenix Security 2025.

AI Institute for Agent-based Cyber Threat Intelligence and Operation

Supported by the National Science Foundation under grant no. IIS-2229876 and is supported in part by funds provided by the NSF, by the Department of Homeland Security, and by IBM.

https://action.ucsb.edu