Securing Dynamic Robotic Coordination in Unpredictable Environments: Enhancing Trust through Adaptive Learning and Cyber Defense

Giselle Roman (California State University, Fullerton)

Advisor: Dr. Yugyung Lee (University of Missouri-Kansas City)

NSF REU AI-Empowered Cybersecurity

ABSTRACT

With the advancement of robotic systems in unpredictable and unsafe environments, significant cybersecurity challenges arise. Attacks can degrade trust and disrupt coordination between robots and humans.

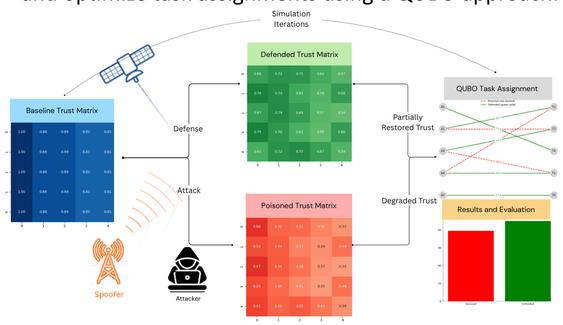
These simulations model adversarial attacks that *poison* trust between agents, reducing effective collaborative task allocation. The approach includes randomized trust degradation and proportional defense countermeasures designed to partially restore trust scores.

To optimize task assignments under trust constraints, we formulated a Quadratic Unconstrained Binary Optimization (QUBO) problem and solved it using a NEAL sampler.

Results show that while adversarial attacks lower trust in task assignment, defense countermeasures can partially restore trust, supporting more effective multi-agent planning in adversarial settings.

METHODS

I modeled a fleet of robots in adversarial settings, simulate trust attacks, apply defense countermeasures, and optimize task assignments using a QUBO approach.



Trust Modeling

- Agents initialized with randomized trust scores.
- Simulates heterogeneous fleet reliability.

Attack Simulation

- Random noise lowers trust scores to mimic cyber-physical threats.
- Models GPS spoofing, communication tampering.

Defense Countermeasures

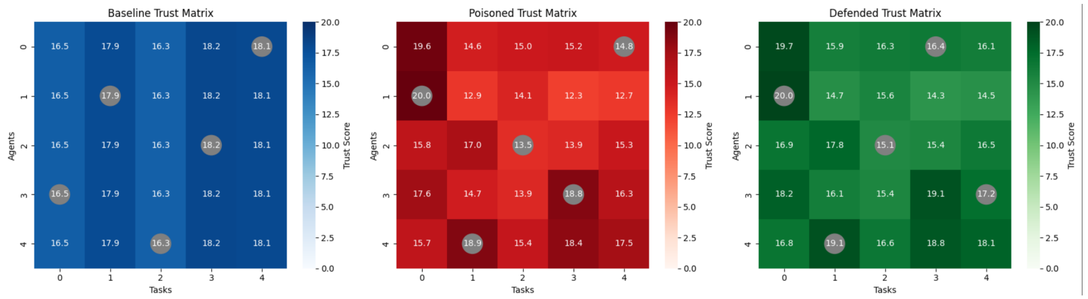
- Proportional restoration based on task success.
- Penalizes repeated failures to encourage reliability.

QUBO Task Assignment

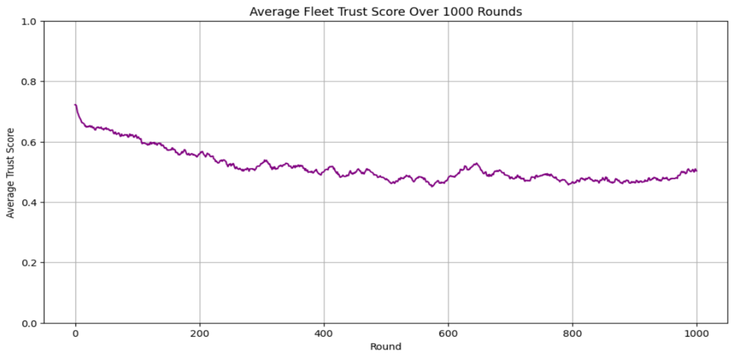
- Formulated to maximize trust-constrained assignments.
- Solved with NEAL sampling to find high-quality solutions.

RESULTS

Simulations evaluate impacts of attacks, defense countermeasures, and optimization on trust-aware robotic coordination.

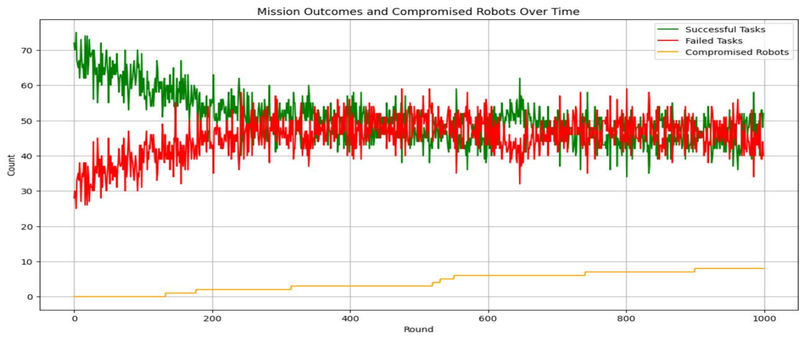


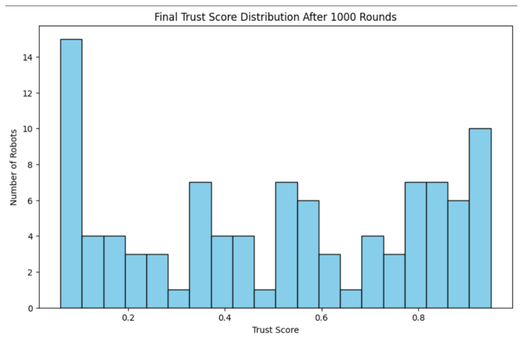
- **Baseline Trust Matrix**: High agent-task confidence before any attack.
- **Poisoned Trust Matrix**: Adversarial attack degrades trust scores across agents.
- **Defended Trust Matrix**: Defense countermeasures partially restore trust, enabling improved assignments.



- Tracks the mean trust level of all robots during repeated attack and defense cycles.
- Attacks cause a steady decline in overall trust, reflecting system vulnerability.
- Defense countermeasures help level off the decline, maintaining trust at moderate levels.
- System can still support coordinated task assignments even under ongoing adversarial conditions.

- Counts of successful tasks (green), failed tasks (red), and compromised robots (yellow) over 1000 rounds.
- Defense reduces compromise spikes, maintains mission throughput.
- Highlights system resilience under repeated adversarial attacks.





- Shows the spread of trust scores across all robots at the end of the simulation.
- Most robots cluster at either low trust (near 0.0) or high trust (near 0.9–1.0), indicating partial recovery but uneven resilience.
- The peak at low trust reflects compromised or unreliable agents that defense could not fully restore.
- The peak at high trust represents agents that maintained reliable behavior under attack and defense cycles.
- Highlights the realistic need for identifying, isolating, or reassigning low-trust units in adversarial environments.

CONCLUSION

- **Adversarial attacks**, such as GPS spoofing or communication interference, systematically degrade trust scores among collaborating robots, leading to poor task allocation and reduced mission success.
- **Simulated defense countermeasures** apply proportional restoration based on agent performance, partially recovering trust and helping maintain reliable assignments even under attack conditions.
- **QUBO task assignment** problem enables optimization under degraded trust, supporting effective agent-task pairings despite uncertainty.
- **Results** demonstrate that integrating trust modeling, anomaly detection, defense strategies, and optimization improves resilience and coordination for robotic fleets in post-disaster or adversarial environments.
- **This approach offers a framework** for adaptive, trust-aware planning that can sustain critical operations despite ongoing cyber-physical threats.

REFERENCES

1. Bojchevski, Aleksandar, and Stephan Günnemann. “Adversarial Attacks on Node Embeddings via Graph Poisoning.” ArXiv.org, 2018, arxiv.org/abs/1809.01093.
2. Haskard, Adam, and Damith Herath. “Secure Robotics: Navigating Challenges at the Nexus of Safety, Trust, and Cybersecurity in Cyber-Physical Systems.” ACM Computing Surveys, Association for Computing Machinery, Mar. 2025, https://doi.org/10.1145/3723050.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation through the REU program at the University of Missouri–Kansas City. I would like to thank Dr. Yugyung Lee for her mentorship, guidance, and valuable feedback throughout this project.