Trust-Aware Task Allocation Under Adversarial Attacks in Multi-Agent Robotic Systems

Giselle Roman (California State University, Fullerton)

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

NSF REU AI-Empowered Cybersecurity

ABSTRACT

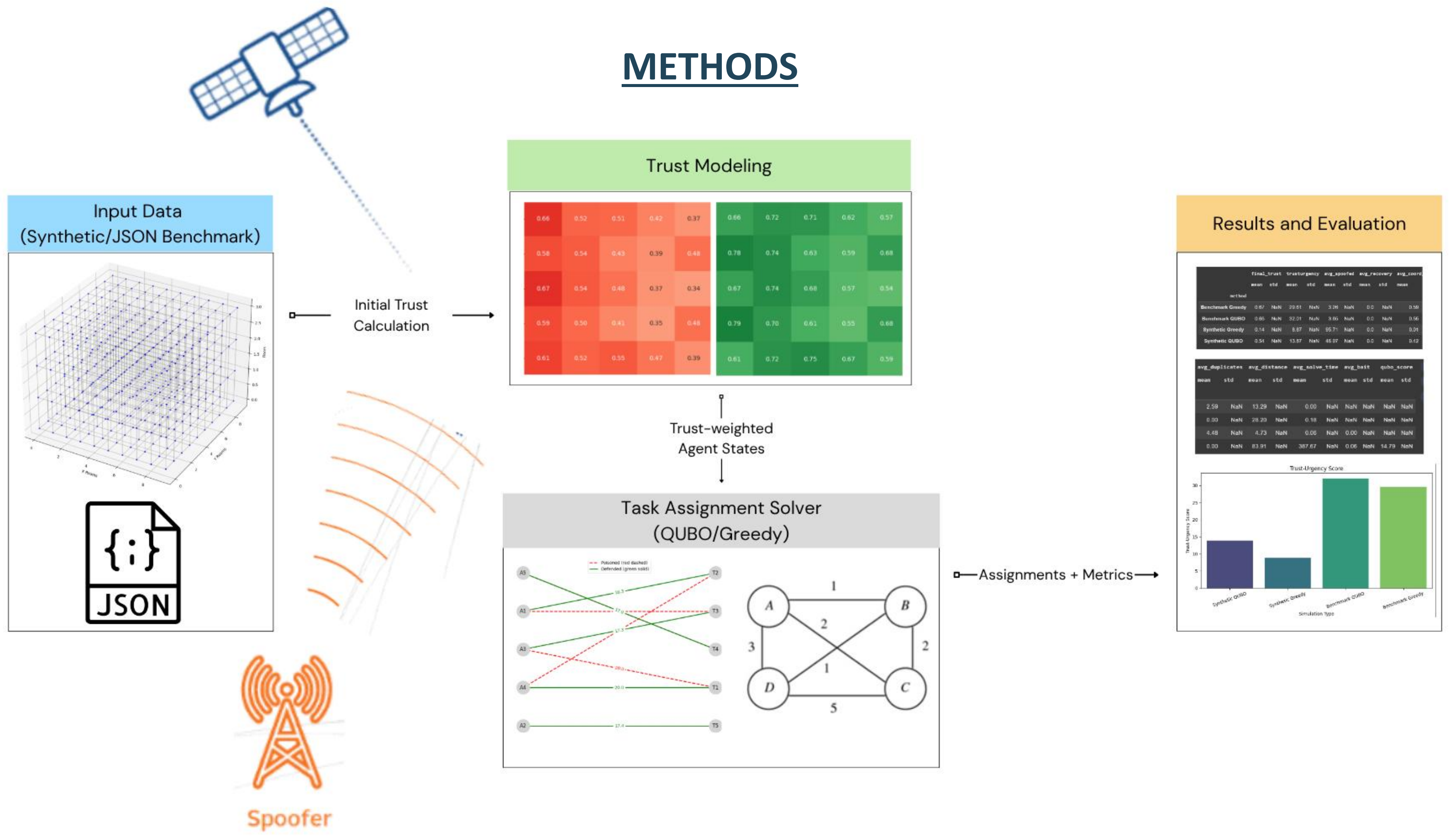
With the advancement of robotic systems in unpredictable and unsafe environments, significant cybersecurity challenges arise. Attacks can degrade trust and disrupt coordination between robots and humans.

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

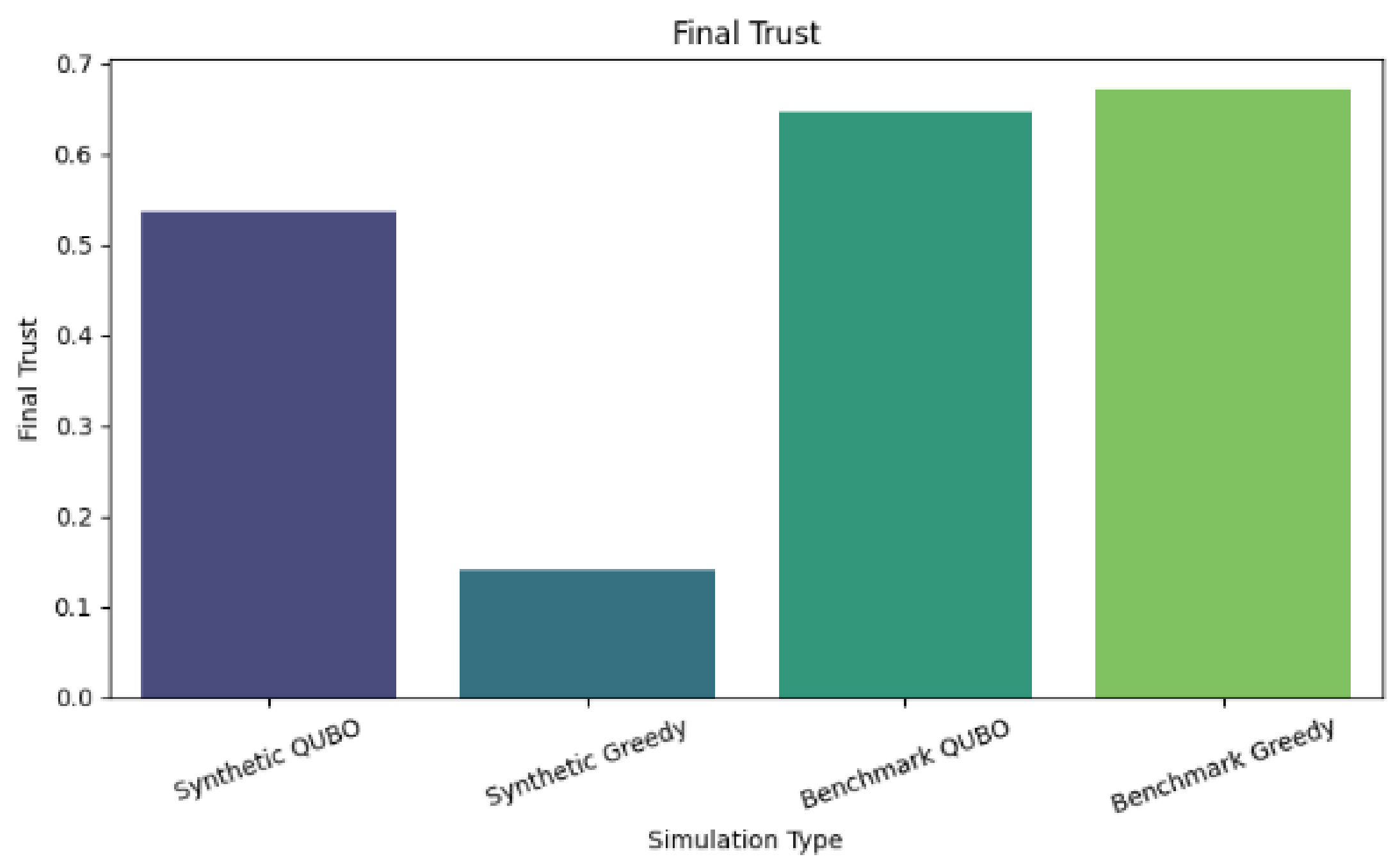

Input Data
(Synthetic/JSON Benchmark)
{;}
JSON
Initial Trust Calculation
Trust Modeling
Trust-weighted Agent States
Task Assignment Solver (QUBO/Greedy)
Assignments + Metrics
Results and Evaluation
Trust-Urgency Score
Spoofer

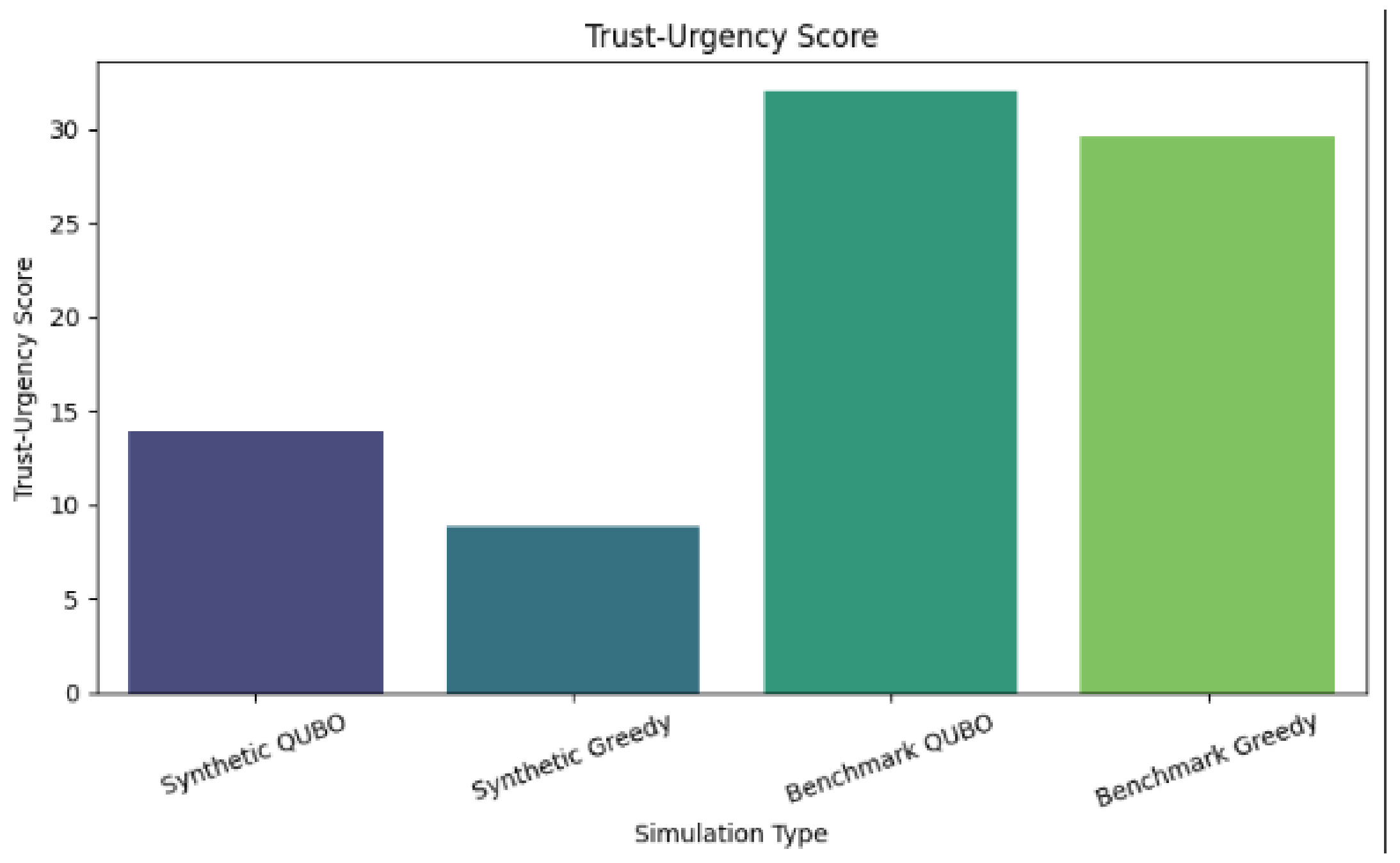
A trust-aware task allocation pipeline was developed for multi-agent robotic systems operating in adversarial environments. Simulations were conducted using both synthetic datasets and standardized JSON benchmark files, each defining agent-task-room layouts, urgency levels, and coordination constraints.

Initial trust values were assigned to each agent based on distance, role, and environmental noise. A spoofing mechanism randomly degraded agent reliability, triggering dynamic trust updates across simulation rounds. Trust scores influenced how agents were prioritized for assignments.

Two solvers were used: a greedy baseline and a QUBO-based optimizer. The greedy solver selected agents based on urgency and proximity. The QUBO solver encoded trust, urgency, synergy, and precedence into a quadratic unconstrained binary optimization model and solved it using a Tabu sampler.
Key metrics tracked across 100 rounds included final trust, trust-urgency score, spoofing rate, coordination success, average distance, duplicate assignments, and solver runtime. Each configuration was repeated over 10 independent runs for statistical robustness.

RESULTS





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.