Title: Secure Robotics: Navigating Challenges at the Nexus of Safety, Trust, and Cybersecurity in Cyber-Physical Systems

Authors: Adam Haskard & Damith Herath

Publication Year: 2025

Key Ideas

The article introduces secure robotics as a unified interdisciplinary framework that integrates trust, safety, and cybersecurity, domains that have traditionally been analyzed in isolation. This approach responds to the growing presence of cyber-physical threats and the increasing need for robotic systems that are resilient, adaptable, and trustworthy. As robots are increasingly deployed in real-world environments, ensuring human safety and adhering to ethical principles become crucial. The authors reference Asimov's laws to stress that robotic systems should align with human values and ethics. The article discusses the importance of integrating societal norms into robotic systems to facilitate smoother interaction within human environments. Trust is presented not as a binary concept but as something that evolves over time. Human-robot interaction (HRI) is shaped by psychological models of trust and risk management, with an emphasis on understanding the contextual factors that affect trust development. In augmented reality systems, where virtual elements are superimposed onto the physical world, unique threats arise, such as sensory spoofing, unauthorized access, and cyberattacks. The authors argue that combining cyber-physical systems (CPS) with the computational capabilities of information systems (IS) enables better interaction with the physical world. A taxonomy of trust-relevant failures in HRI is provided, including design failures, system failures, expectation mismatches, and user errors, each influencing trust differently.

Contributions

This article contributes by defining secure robotics as a new interdisciplinary field that merges trust, safety, and cybersecurity within CPS. It outlines the cybersecurity challenges and vulnerabilities that robots face, especially when trust and safety intersect. Challenges such as weak data integrity, insecure networks, and ineffective human-machine collaboration are linked to poor communication and weak authentication. To enhance system reliability, the authors suggest improving transparency, clearly communicating intent, and increasing dependability. The article highlights threats to human-robot trust caused by cyberattacks like denial-of-service, dynamic manipulation, and replay attacks. It proposes a conceptual framework to understand and mitigate trust failures in robotic systems, particularly those involving unauthorized access, data manipulation, or risks of physical harm.

Limitations

Since this unified framework is relatively new, it remains theoretical and lacks experimental validation. The article provides conceptual models but does not offer hands-on implementations or real-world testing. While it raises cybersecurity concerns, it does not go into detail about specific defense tools or system-level implementations. Furthermore, the article does not provide concrete guidelines or alternatives for addressing ethical concerns in robotic system development.

Extend/Improvement

To build on this article, my research will focus on implementing security-aware learning systems for adaptive robots operating in unpredictable environments. By integrating trust and cybersecurity during the learning process, I aim to reinforce learning pipelines, such as imitation or reinforcement learning, where trustworthiness is treated as a core goal. Drawing on the taxonomy of failures presented in the article, I plan to design an adaptive feedback system that evaluates and addresses communication gaps affecting user trust during the robot's learning phase.