Birds of a Feather (CPS Leaders and CPS Rising Stars)

Moderator: Edward Lee (UC Berkeley)
Scribe: Fanxin Kong (University of Notre Dame)

Participants: David Corman (NSF-Retired), Meiyi Ma (Vanderbilt), Miroslav Pajic (Duke), Patrick Borges-Rodrigues (NJIT), Sayan Mitra (UIUC), Zejiang Wang (University of Texas-Dallas), Zixiao Ma (Binghamton University)

Physical Guardrail in CPS as an Enabler for Leveraging AI/SI

Naïve Physical Guardrail: Apply the Brakes on Any Doubt



Edward: theme of the session

- Senior researchers and rising stars
- Identify CPS mission, challenges, risks and most important research areas
- Help define NSF CPS (could be next proposal, ideas)

Self introduction

- Patrick Borges-Rodrigues (NJIT)
  - Understanding facility using CPS, e.g,, human working with robots
  - Making people trust these systems and cooperate
- Zejiang Wang (University of Texas-Dallas)
  - Control and optimization for ground vehicles and transportation
  - Cost-aware control design
- Zixiao Ma (Binghamton University)
  - Privacy-reserved power load control/management for power grid
- Sayan Mitra (UIUC)
  - Formal verification of CPS, recent focus on robotics (e.g., air vehicles)
  - Moonshot problems from CPS perspective

Self introduction

- David Corman (NSF-Retired),
  - Big problems today with CPS flavor
  - Safety in the era of AI, what is the fundamental meaning, real safety definition in AI and autonomy
  - Resilience in CPS, e.g., environment (resilience against weather)
- Meiyi Ma (Vanderbilt)
  - Trust and explainable AI,
  - Co-design between community and tech
  - Explainability in CPS: AI explainability is not enough. We need user-understandable system design.
- Fanxin Kong (Notre Dame)
  - Generalizability: base model that can integrate common knowledge of CPS
  - Adaptability: evolving with new tasks, new data, new environment….
  - Low SWaP: small size, light weight, and low power AI for CPS

Challenges

- Trust between CPS and human
  - Make people trust CPS systems and cooperate
- Explainable CPS
  - Beed user-understandable system design
- Generalizability
  - Base model that can integrate common knowledge of CPS
- Adaptability
  - Evolving with new tasks, new data, new environment….
- Low SWaP
  - Small size, light weight, and low power AI/SI for CPS

Proposal: Physical Guardrail in CPS as an Enabler for Leveraging AI/SI

- Physical guardrail in CPS
  - Building trust among human, SI, and the system
  - History: cyber-physical contracts, e.g., flight envelop protection
  - Related to safety and resilience
- Hierarchical design
  - Need multiple levels of guardrails from cyber to physical layers, from perception to actuation….
  - A set of rules not enough, how to go beyond it
- Explainable-enough CPS
  - Explainability enough to earn users' trust while not reveal too deep details
- Guardrail from CPS perspective
  - New definition, new design, deployment to various applications
- Iteration between designing guardrails and improving AI/SI models
  - AI/SI is probabilistic and may trigger guardrails often at runtime
  - Need to co-design guardrail and AI/SI model to alleviate this