AI/Machine Learning & Cyber-Physical Systems

K. Khorasani, Ph.D., P. Eng.
Professor
Department of Electrical & Computer Engineering
Concordia Institute of Aerospace Design and Innovation
Concordia University
Montreal, Quebec Canada

September 19, 2019

*** Disclaimer: Some of the figures included in this presentation do not belong to the presenter and are "borrowed" from the web for the purpose of this talk. Exact citations can be provided upon request.**

Research Motivations in CPS

- Advances in control, computing, and communications have led to development of highly interconnected, computer networked, and distributed systems that is commonly known as cyber-physical systems (CPS).

Applications:

- **Aerospace/aviation and next generation transportation systems**
- **Process and automation systems (Oil & Gas, Chemical)**
- **Smart Grid System; Smart Cities**
- **Critical infrastructure monitoring (Dams& Bridges)**

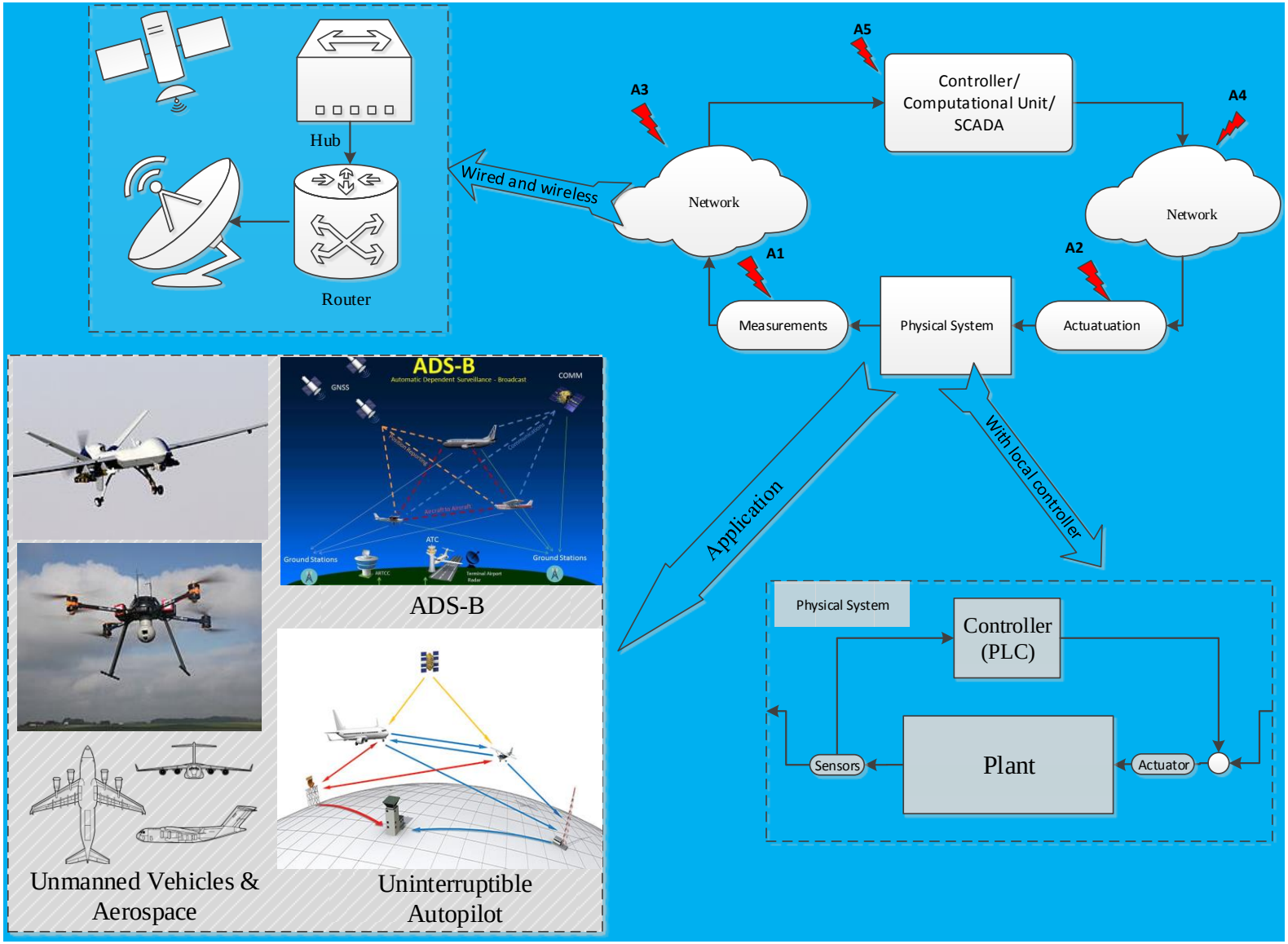
Objectives and Challenges

- Develop *distributed* and cooperative strategies for *optimal allocation & distribution of information and resources*.

- Subject to bandwidth, latencies, real-time, computing, and energy constraints by using a ***hybrid framework*** through optimally integrating model-based *with* machine learning (ML) or computationally intelligent (CI) and AI-based theories.

- Challenges:
  - **Complexity: Cyber-physical systems (CPS); System of systems.**
  - **Communication: Limited bandwidth and connectivity. What? When? To whom?**
  - **Arbitration: Multi-objective optimization problem. Team vs. Individual goals.**
  - **Computational/asset/vehicle resources: Is always limited.**
  - **Cyber-security, assurance, trustworthiness, confidence, and guaranteed reliability.**

CPS Security Challenges



Research Challenges in Cyber-Physical Systems using AI

- Unfortunately, due to ambiguities that arise from uncertainties, adversarial cyberattacks, human operators will lose trust in these systems and this could lead to hesitation/doubts by operators using these assets in day-to-day missions.

- Of paramount importance is to enable human operators to make distinctions between *safety critical decisions* versus *critical functionalities* without compromising safety and trustworthiness of the assets operations.

- Novel methodologies are required to certify and assure CI/ML-based assets as being trustworthy.

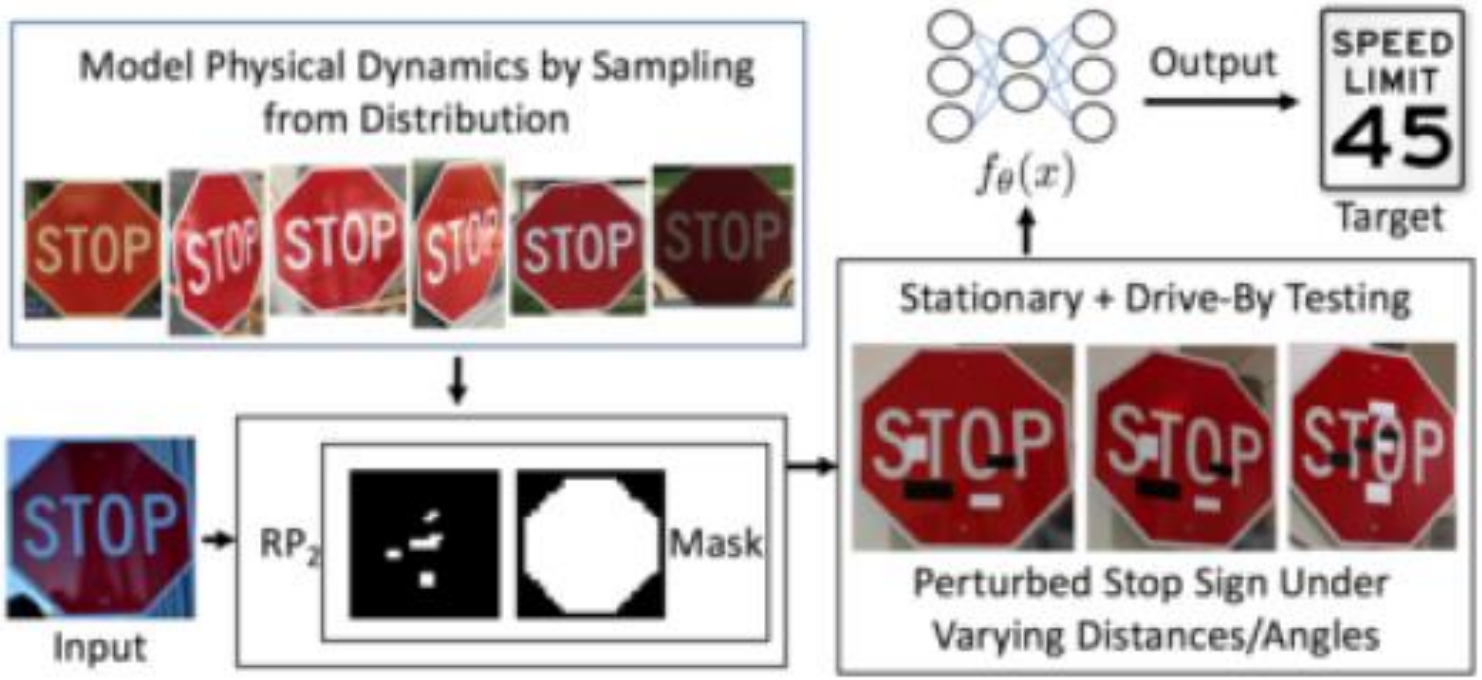
- *Assurance of safety* and *functional correctness* of algorithms developed by CI/ML solutions play crucial roles in ensuring wider acceptance and their utilization by C&C operators.

Research Challenges in Cyber-Physical Systems using AI

- Other challenges related to CI/ML/AI-based trustworthiness, assurance, and confidence assessments can be accomplished through:
  - *Model validation* that determines if the model used corresponds to reliable representations of the actual operating environment,
  - *Evaluation of the expected outcomes* that determine if allocation of rewards under a given policy will yield the specified outcomes,
  - *Operator interpretation* that determines if the autonomous decision making process captures the operator intentions and translates them into suitable tasks, and
  - *Historical performance evaluation* that determines efficiency and efficacy of previous and scenarios associated with the same or similar problems.

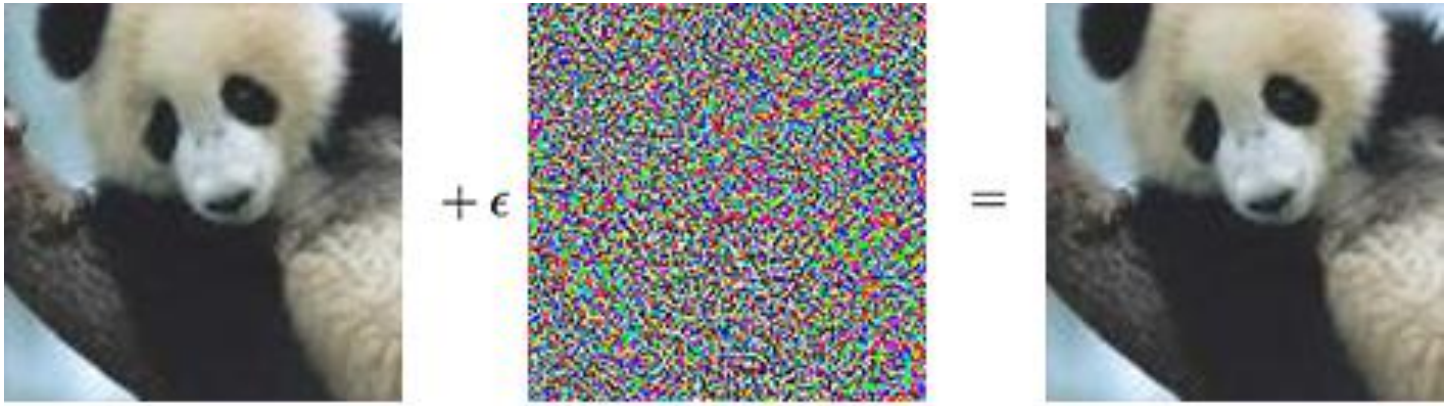
Adversarial Example – One Challenge in AI-based Techniques

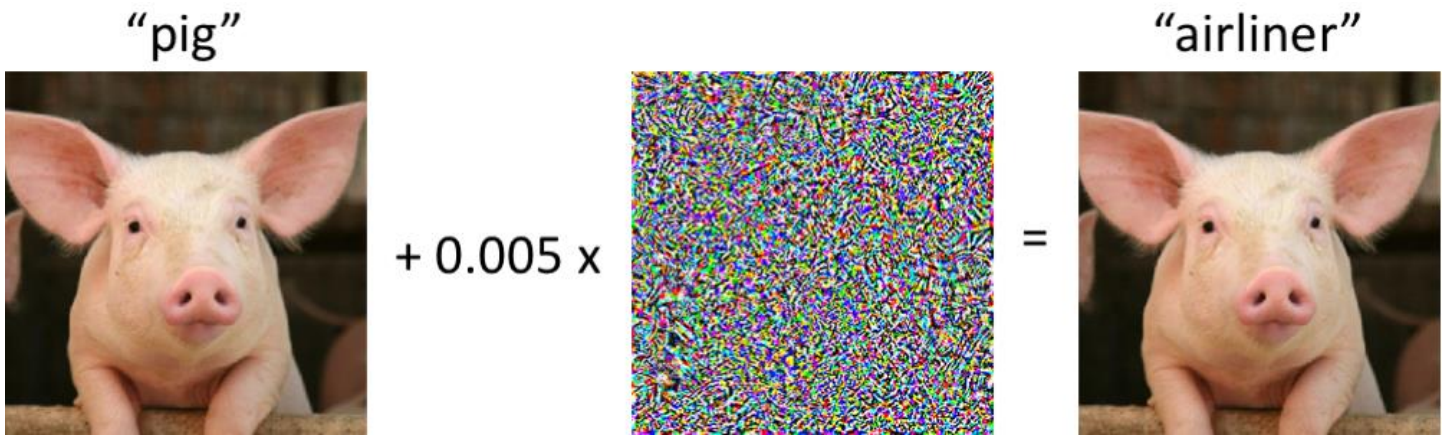
- Adversarial Example (AE) is an intentionally crafted input channel of a Deep NN that results in a significant impact (and incorrectness) in output of the network.
- Need investigation to fully understand when and under what conditions a trained Deep NN is subjected to an AE.



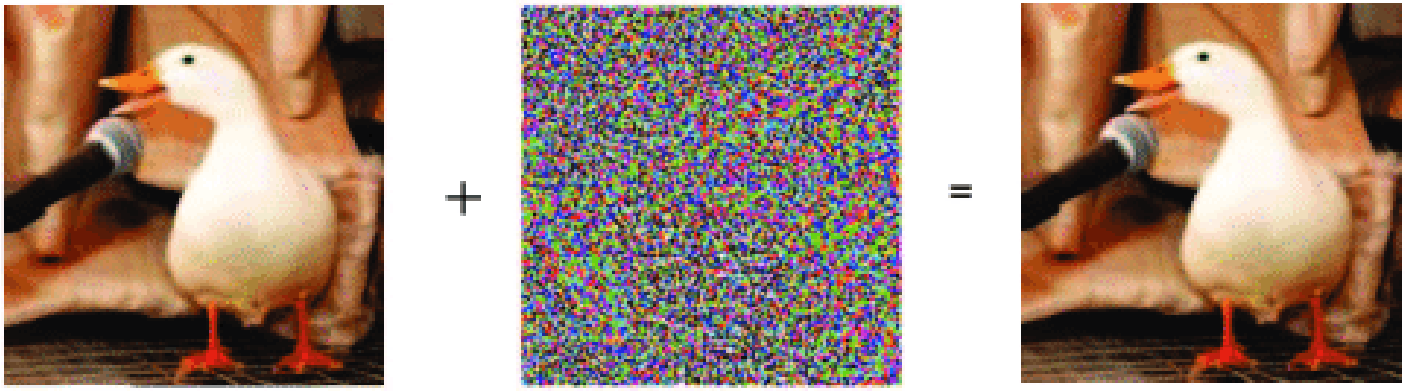
Eukholt, et. al. – Robust Physical-World Attacks on Deep Learning Visual Classification, 2018. A perturbed stop sign is miss-classified as a SPEED LIMIT.

Adversarial Example – One Challenge in AI-based Techniques





Adversarial Example – One Challenge in AI-based Techniques



Research Challenges in Cyber-Physical Systems using AI

- How to develop a set of public policy standards and frameworks for autonomous (including UAVs), AI, and CPS such that their explainability, interpretability, trust, safety and reliability can be measured, assured and their safety and reliability guaranteed.

- In particular, how to develop alternative approaches to AI that are explainable, and therefore certifiable and trustworthy.

- What principles of equality, inclusiveness, and non-discriminative practices should be proposed to ensure and preserve fairness, human dignity and rights that are central to economic and social impacts of proposed technology policies and legal regulations.

- What are the key pillars of effective policies and regulations to address the challenges related to transparency, predictability, assurance, and accountability of the autonomous (including UAVs), AI, and CPS?

Imagine a future…