Artificial Intelligence and Machine Learning in Autonomic and Automated Security

Aron Laszka
University of Houston

Computational Cybersecurity in Compromised Environments
2021 Fall Workshop | October 27-28 | Virtual

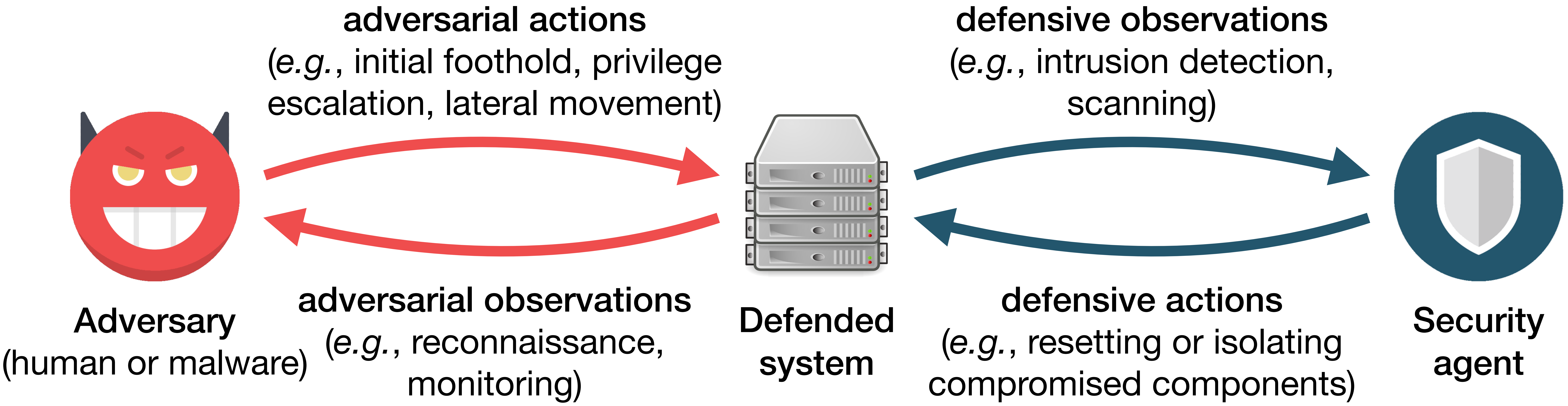
Motivation

Autonomic and Automated Security

- Nowadays, security practitioners face a **continuously and rapidly evolving threat landscape**, and must continuously adopt novel defense techniques
- However, determined and resourceful adversaries **may breach even such well-protected systems**; hence, it behooves defenders to consider cyber-risk management beyond the first lines of defense
- To minimize the impact of security breaches, incident responders must **promptly and effectively mitigate detected intrusions** (*e.g.*, isolate and reset compromised components)
- Since human decision making can be slow and error-prone, especially in complex and uncertain environments, responders must be supported by **autonomic and automated security tools**

Vision

Artificial Intelligence based Security Agents



- Model the cybersecurity **conflict as a multi-agent partially-observable Markov decision process**
- Find optimal defense using **multi-agent deep reinforcement learning**

Exemplary Problem Domains

Project Scope

Cyber-Attack Resilient Control

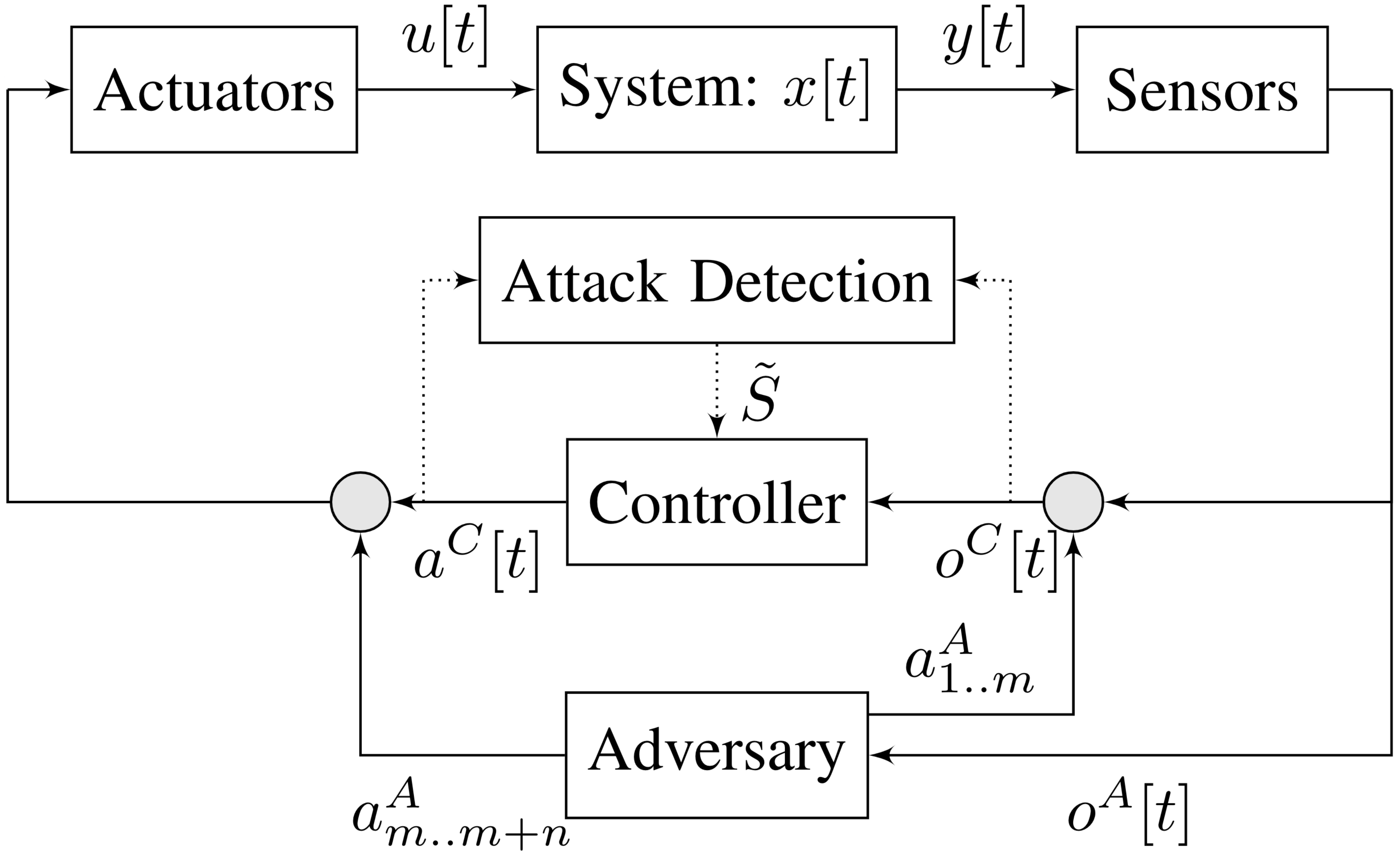
- by **compromising and tampering with cyber-physical systems** (CPS), such as critical infrastructure (*e.g.*, power systems), adversaries may inflict financial losses, physical damage, and even bodily harm
- we consider mitigating such attacks by **adjusting the control policies of the systems** to compensate for adversarial tampering

Strategic Remote Attestation of Internet of Things (IoT) Devices

- **IoT devices and application** can have significant vulnerabilities
- an important approach for mitigating the impact of exploiting such vulnerabilities is **remote attestation**
- we consider **finding optimal policies for applying remote attestation**, which minimize both cyber risks and computational costs

Mitigation in Cyber-Physical Control Systems

Cyber-Attack Resilient Control



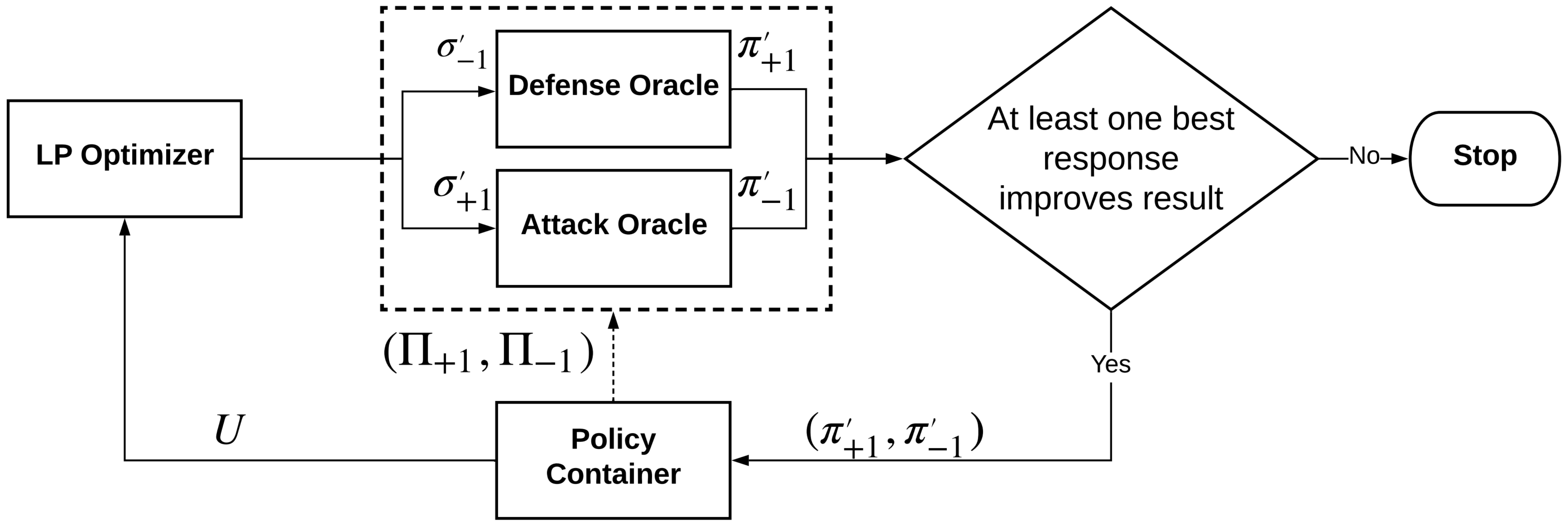
- Adversary has compromised some sensors S_y and actuators S_u and may tamper with their signals a^A
- Defender has detected the intrusion $\tilde{S}$ and mitigates it by changing the controller's policy (*i.e.*, mapping from sensor value o^C to actuation signal a^C)
- Defender's loss is deviation from desired system state (x - $\tilde{x}$)

- *Objective:* find a **resilient control policy** that minimizes the impact of the cyber attack

Solution Approach

Computing Attack-Resilient Control Policies

- Assumption of malicious adversary (*i.e.*, adversary's goal is to maximize the defender's loss)
  → resilient policy is a **Nash equilibrium** of adversarial and defensive policies



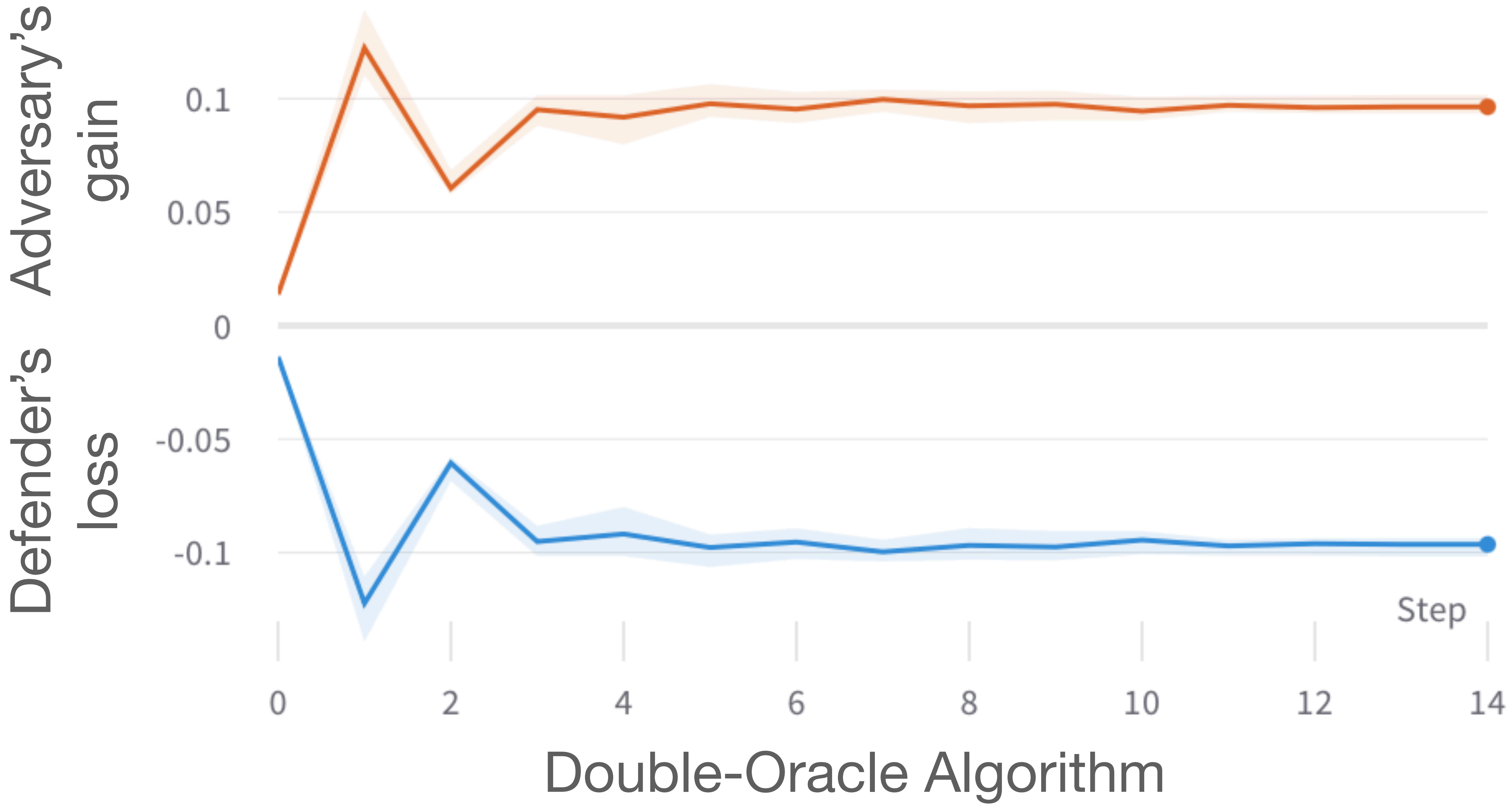
- **Double-Oracle Algorithm**:
  finds an equilibrium by **iteratively computing best-response policies** for the adversary and defender (against equilibrium of prior policies)

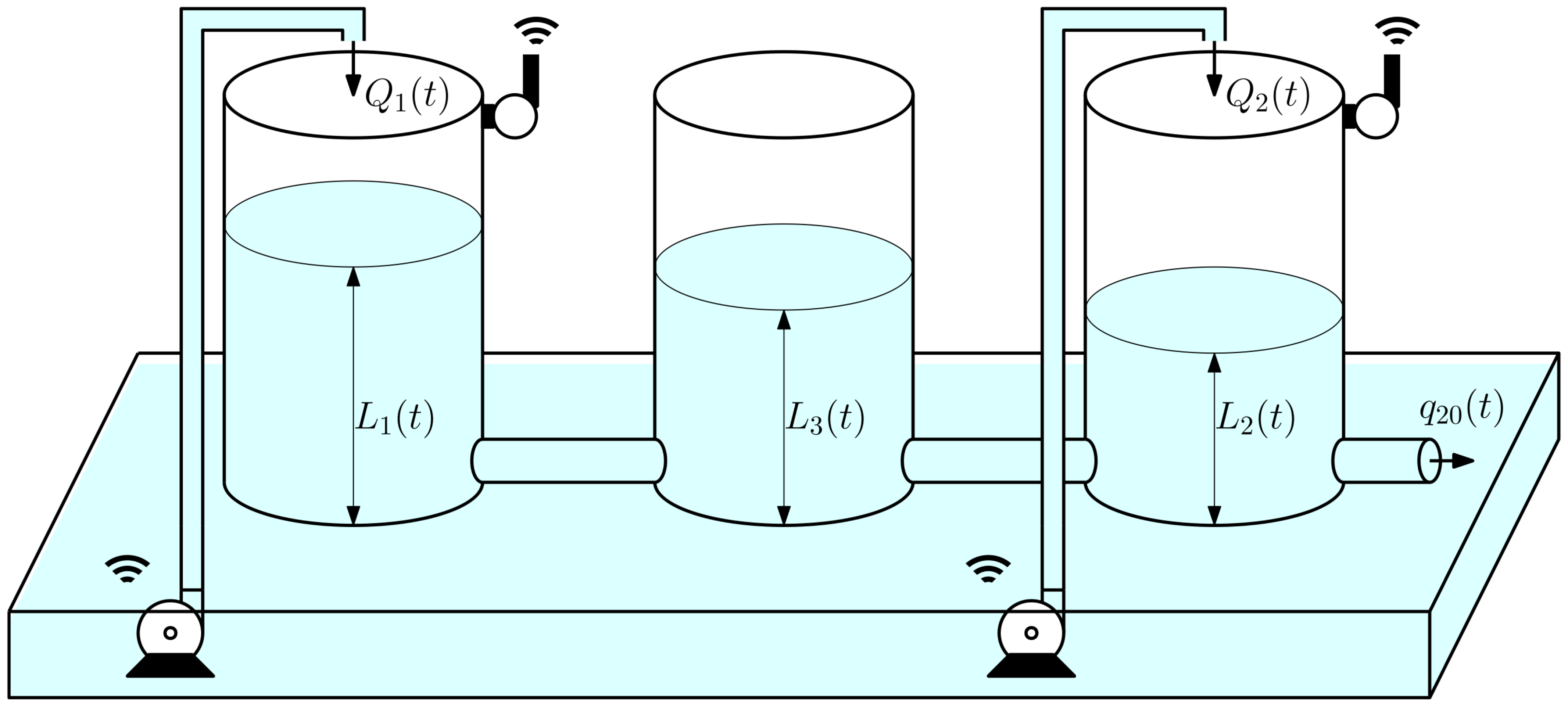
- **Deep Reinforcement Learning**:
  finds a best-response policy against the opponent's policy
  - **deep deterministic policy gradients** (DDPG) in our experiments

Experimental Evaluation

Case Studies

- *Case-study systems:* a **bioreactor** and a system of **three coupled tanks**
- Evolution of the **adversary's gain and defender's loss** over iterations of the double-oracle algorithm **in the three-tanks system:**





Results and Conclusion

Key Findings

- Experimental results show **26% reduction in the defender's loss** in the three-tanks systems (see previous figure)
- Similar results in the bioreactor system
- Resilient control policy can be **trained in a few hours** on commodity hardware
- Control policy is trained to **mitigate a wide range of attack scenarios** (with negligible cost of execution once the policy is trained)
- *Conclusion:* proposed approach is **computationally feasible** and **effective at mitigating cyber-attacks** in cyber-physical control systems

Ongoing and Future Work

Extending the Framework

- *Ongoing work:* adapt framework to **remote attestation of IoT devices**
  - extends the spectrum of problems: cyber-attack resilient control has real-time constraints, continuous actions, and industrial applications, while remote attestation has discrete actions and includes home-user applications
- *Future work:*
  - generalize framework to incorporate **observations beyond detection**
  - incorporate **more cyber actions** (*e.g.*, isolation at the networking level)

Thank you for your attention!
Questions?

Contact Aron Laszka
https://aronlaszka.com/
alaszka@uh.edu