Future of Cyber-Physical Systems (security/resilience)

Saman Zonouz

Associate Professor, Georgia Tech
School of Cybersecurity and Privacy (SCP)
School of Electrical and Computer Engineering (ECE)

Future of CPS Workshop, NSF CPS PI Meeting 2022

CPSec: Cyber-Physical Systems Security Lab

Outline (focused on resilience/security)

CPSec: Cyber-Physical Systems Security Lab

- Important research challenges
- Exciting opportunities for CPS research
- Lessons learned from the past
- Ideas for tech-transfer initiatives

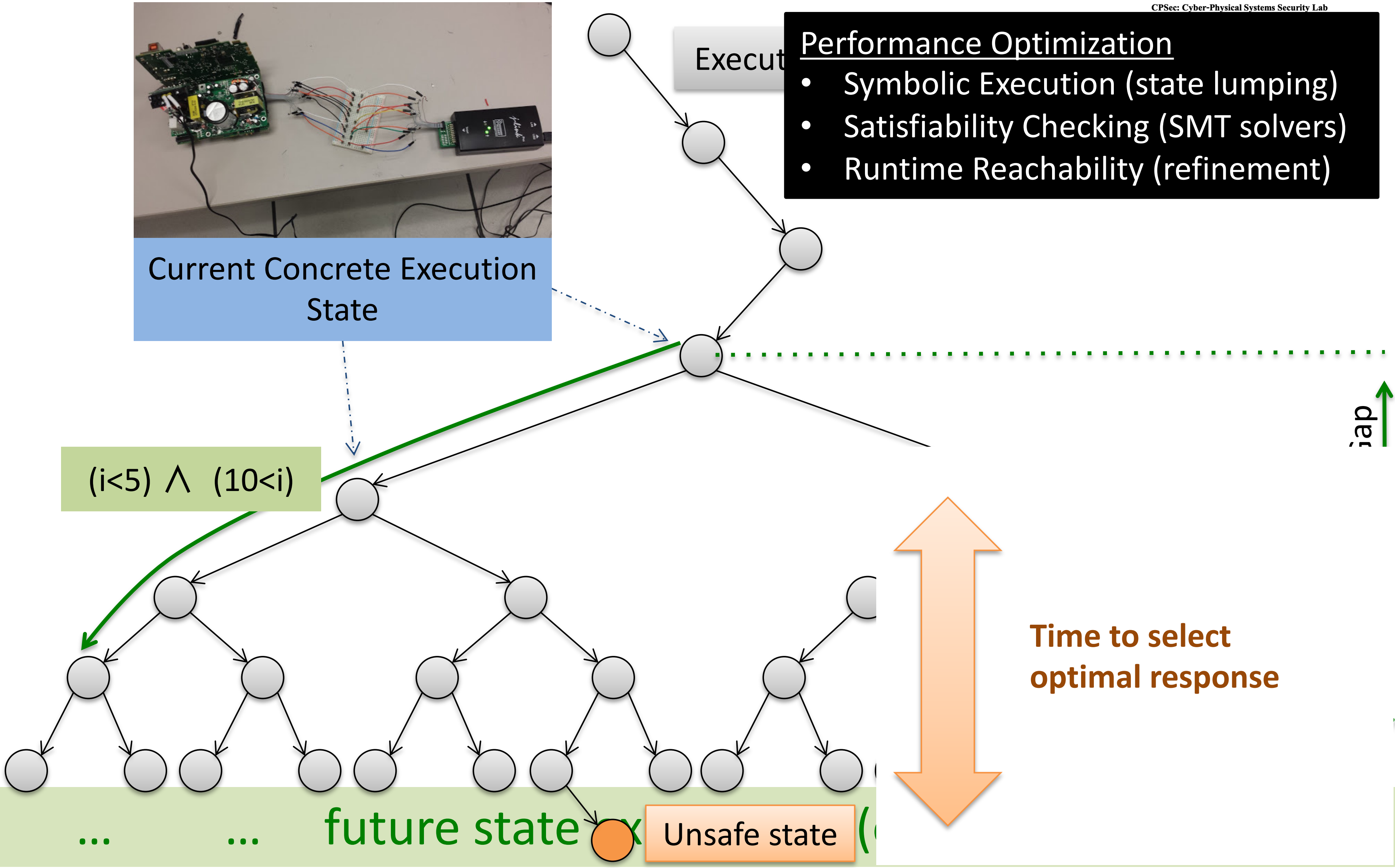
Predictive Situational Awareness

- Online monitoring of the CPS operation to identify potential cybersecurity incidents
- Extensive work on transitioning IT-like real-time monitoring solutions to CPS domain (e.g., mount IMUs to monitor the motion)
- Not always useful in practice due to physics momentum and inertia – chase.com vs Tesla
- **"Ahead-of-Time alerts"** are required to provide time for decision-making on response action selection and its enforcement (potentially in physical components - time-consuming)

JAT Verification [NDSS, ACSAC]

Performance Optimization

- Symbolic Execution (state lumping)
- Satisfiability Checking (SMT solvers)
- Runtime Reachability (refinement)



<u>Physics-Aware</u> Software Analysis

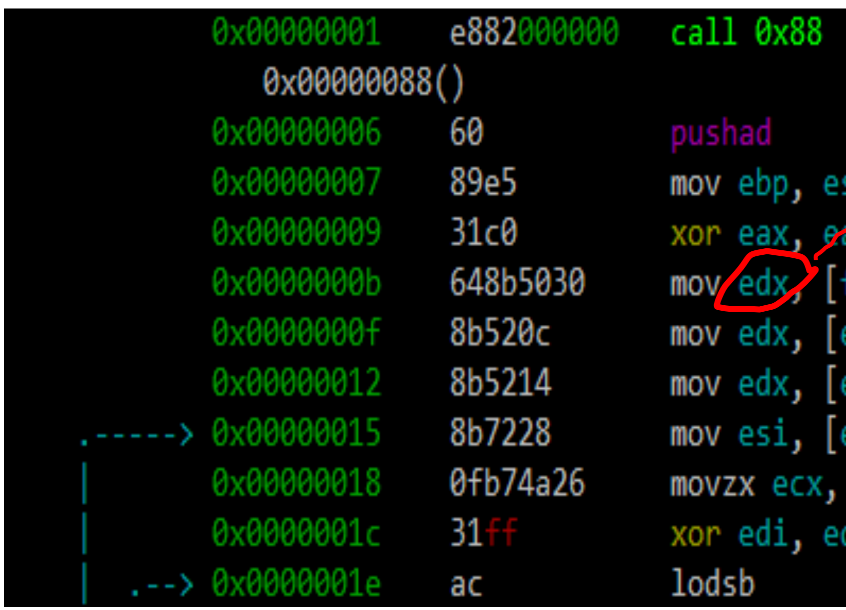
CPSec: Cyber-Physical Systems Security Lab

- Semantic gap (disconnect) between software concepts and physical process concepts
- Nowadays, software analysis tools completely ignore underlying physical dynamics
  - *reverse engineering, vulnerability assessment, hardening (e.g., patching, CFI)*
- All algorithmic vulnerabilities are overlooked
  - *as opposed to conventional SW vuls (UAF, BoF, …)*
- The potential safety consequences of individual SW vulnerabilities are unknown
  - *similarly for attackers, "what value should I overwrite following a heap overflow exploitation?"*

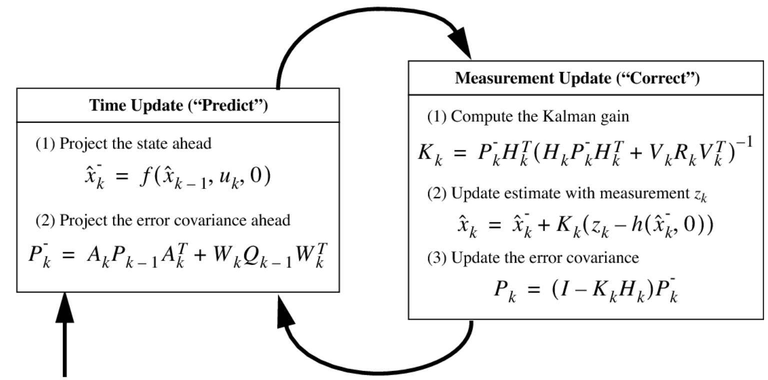
Reversing Control Semantics [MobiSys, DSN]

CPSec: Cyber-Physical Systems Security Lab

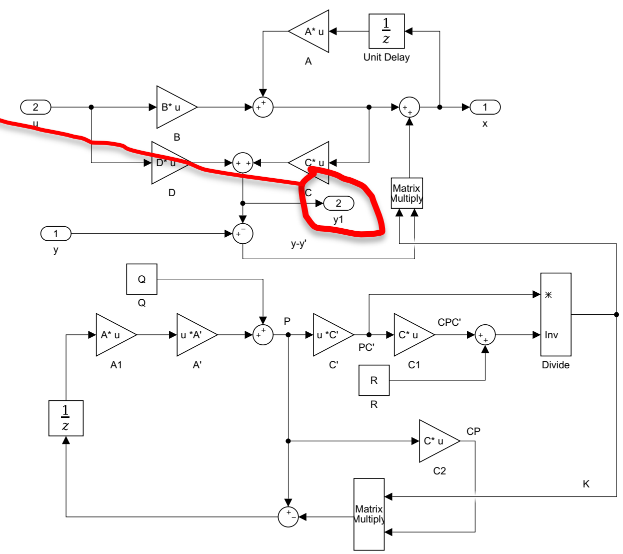
PLC Controller



Low-Level Disassembled Binary Code



?



<u>Human-Assisted</u> Intrusion Response

CPSec: Cyber-Physical Systems Security Lab

- Existing CPS security focuses on prevention (hardening) and monitoring (attack detection)
  - *almost no emphasis on cyber-physical R&R*
- Fully automated R&R is too complex
  - *selection of optimal response policies including both cyber and physical actuation is even harder*
- Promising solutions (e.g., SIEMs) to enable operators to make correct decisions (outage management)
- Next step: human-assisted R&R capabilities
  - *provide operators with a list of 'relevant' potential R&R countermeasures for confirmation*
  - *learning (cost functions) by observing operators passively to imitate them later actively*

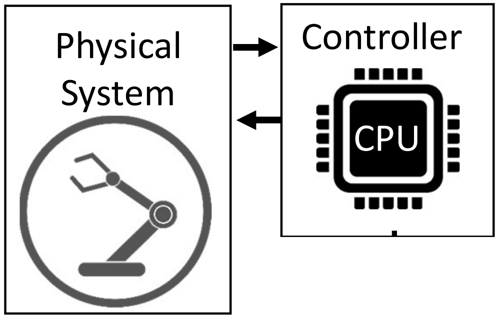
Domain-Specific AI for Security

CPSec: Cyber-Physical Systems Security Lab

- Almost all AI models are optimized for computer vision (e.g., ImageNet competitions)
  - *not always tuned for non-image process/software data*
- Often used blindly for security purposes
  - *process data anomaly/attack detection, binary decompilation, code similarity (bug discovery)*
- Not serving domain-specific requirements
  - *testing data could/should come from a maliciously-designed different attack – lack of robustness*
  - *e.g., sys-wide anomaly detection w/o diagnostics*
- Robustness is a more difficult problem in security
  - *malicious players involved with different attack vectors*

AI-Powered Side Channel Analysis

[CCS]



+ No interference with real-time control

+ Air-gapped detection trusted computing base

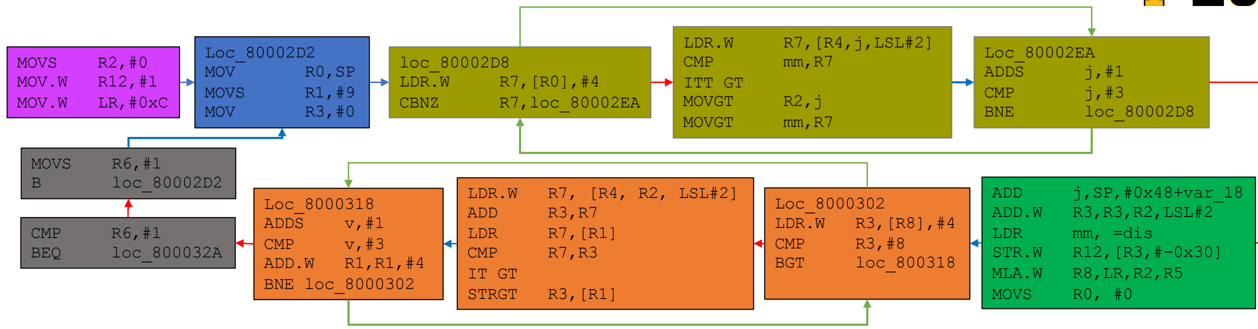
+ Hard to mislead due to tamperproof physics laws that generate side signals

[1] Genkin, et al. "ECDSA key extraction from mobile devices via nonintrusive physical side channels." CCS 2016.
[2] Nazari, et al. "Eddie: Em-based detection of deviations in program execution." ISCA 2017.

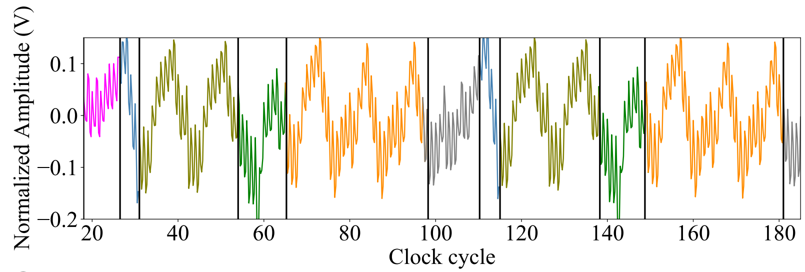
Robustness Against Mimicry Attacks?

[USENIX-Sec]

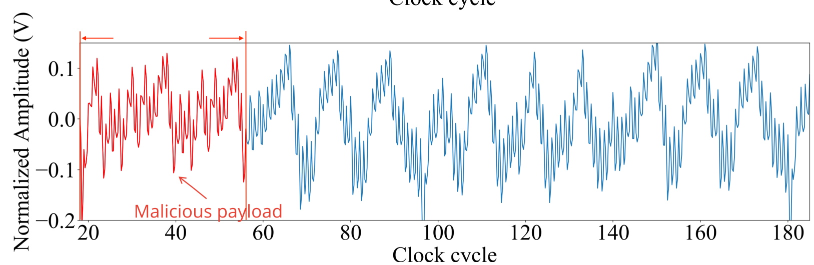
Controller program

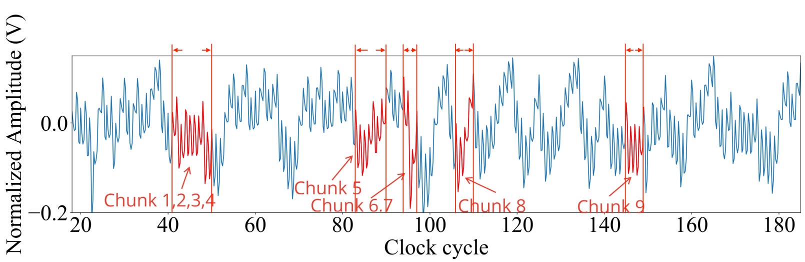


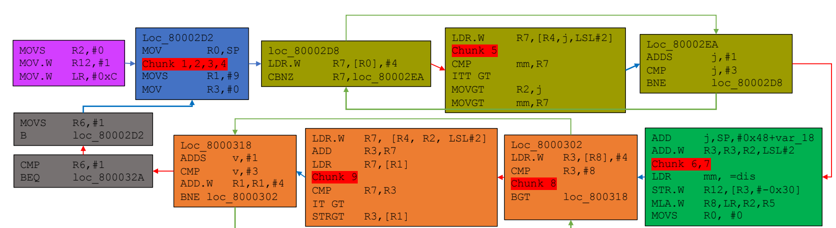
Power signal



Trivial malware injection (**Detected**)







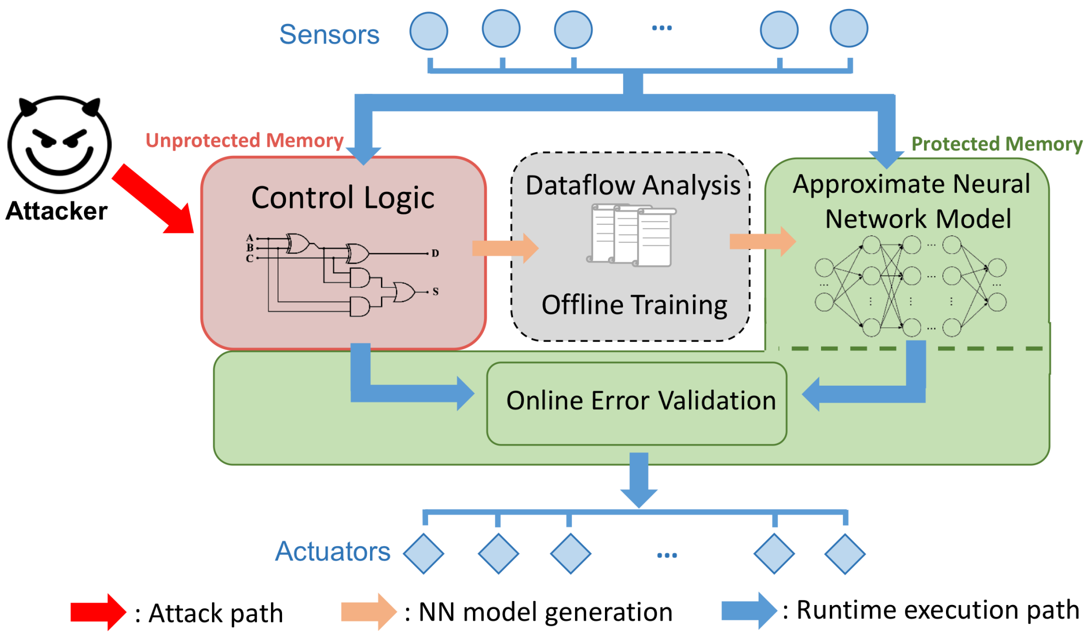
Optimal Chunked Malware Injection (NOT Detected)

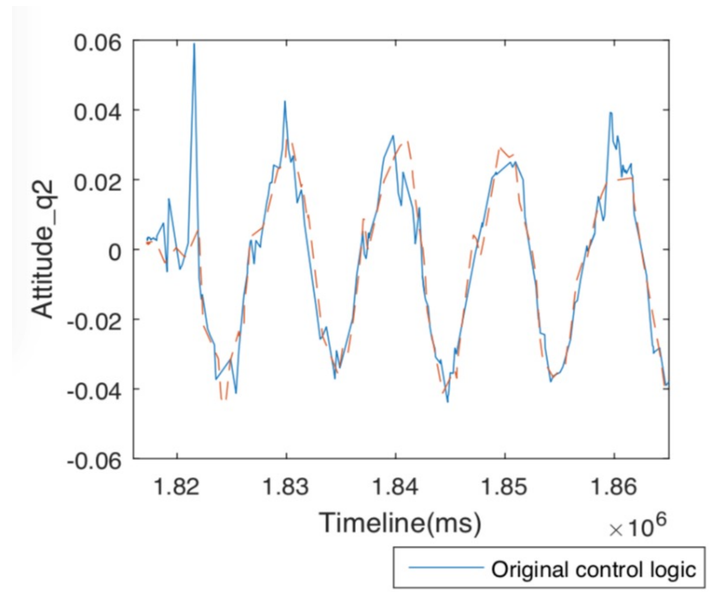
Trustworthiness w/ Untrusted (edge) AI

- AI solutions are getting more complicated
  - *e.g., in terms of DNN size, architectural complexity*
- "Verified AI" for real-world large models could take time to be practical (industry reluctance)
  - *similar to SW verification efforts – code bases get more complex while verification solutions improve*
- Edge AI for the communication-computation tradeoff
  - *less secure (e.g., due to security support/DEP in MCUs)*
- Ensure safety for systems including AI modules, which may act wildly
  - *top-down system-wide (to detect/ignore suspicious AI)*
- Security-oriented DNN debloating/pruning [NeurIPS'21]
  - *to simplify verification at the cost of suboptimal control*
  - *create a verifiable suboptimal small replica (surrogate) of the main optimal controller – used for safety monitoring and response*

DNN-based Surrogate for Assurance

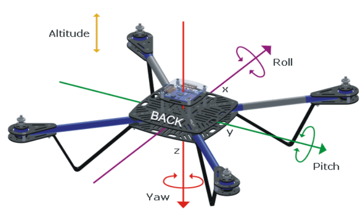
[NDSS, RAID, NeurIPS]





$$x(n+1) = f(x(n), u(n)) + w(n)$$

$$y(n) = h(x(n)) + v(n)$$



Conclusion

- Predictive Situational Awareness
- Physics-Aware Software Analysis
- Human-Assisted Intrusion Response
- Domain-Specific AI for Security
- Trustworthiness w/ Untrusted (edge) AI

CPSec: Cyber-Physical Systems Security Lab

Positions available in
Trustworthy ML and CPS security

Saman Zonouz saman.zonouz@gatech.edu

Thank You!