KTH ROYAL INSTITUTE
OF TECHNOLOGY

Secure Networked Control Systems: Closing the Loop over Malicious Networks

Henrik Sandberg (hsan@kth.se)

Division of Decision and Control Systems, EECS, KTH

Acknowledgements

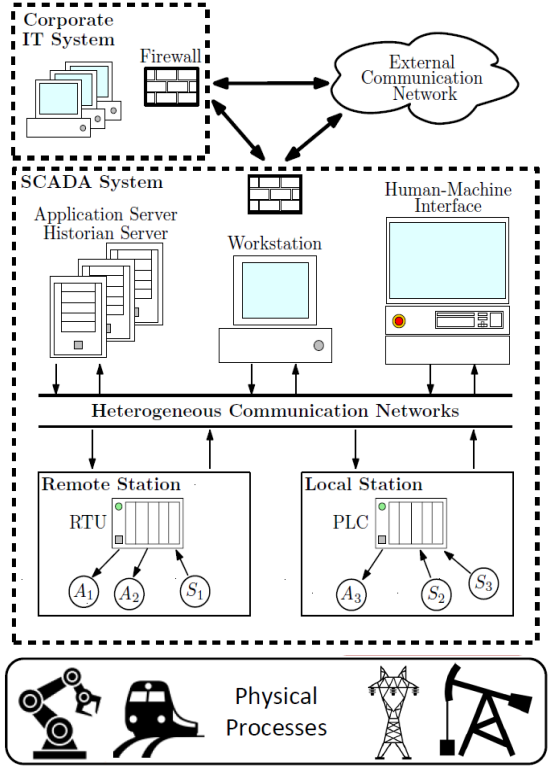
- **KTH:** David Umsonst, Jezdimir Milosevic, Kaveh Paridari, György Dán, Karl Henrik Johansson
- **Uppsala University:** André Teixeira
- **Dell-EMC:** Niamh O'Mahony
- **UTRC:** Alie El-Din Mady, Rohan Chabukswar, Menouer Boubekeur
- **Sponsors:** Myndigheten för samhällsskydd och beredskap, digital futures, Smart Energy Systems ERA-Net, Vetenskapsrådet, Energimyndigheten, SSF

Outline

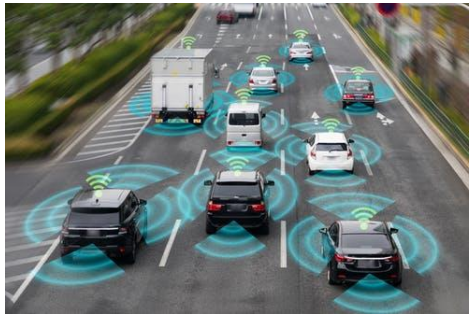
- Security challenges for industrial control systems
- CPS security in networked control
- Work in our group
  - Energy management systems, stealthy attacks, and security metrics
  - Confidentiality in the closed loop
- Summary and outlook

Cyber-Physical Systems

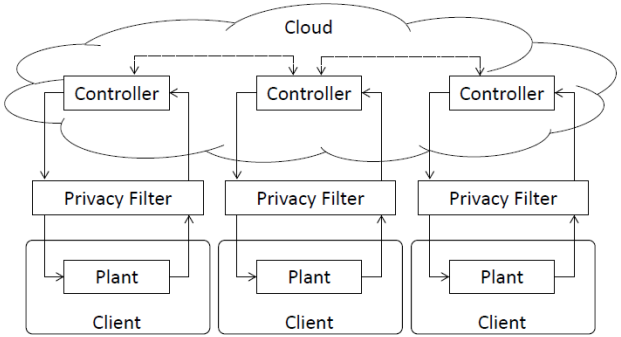
Industrial Control System (ICS)



Autonomous vehicles



Cloud-based Control and IoT



Example 1: The Stuxnet Worm (2010)

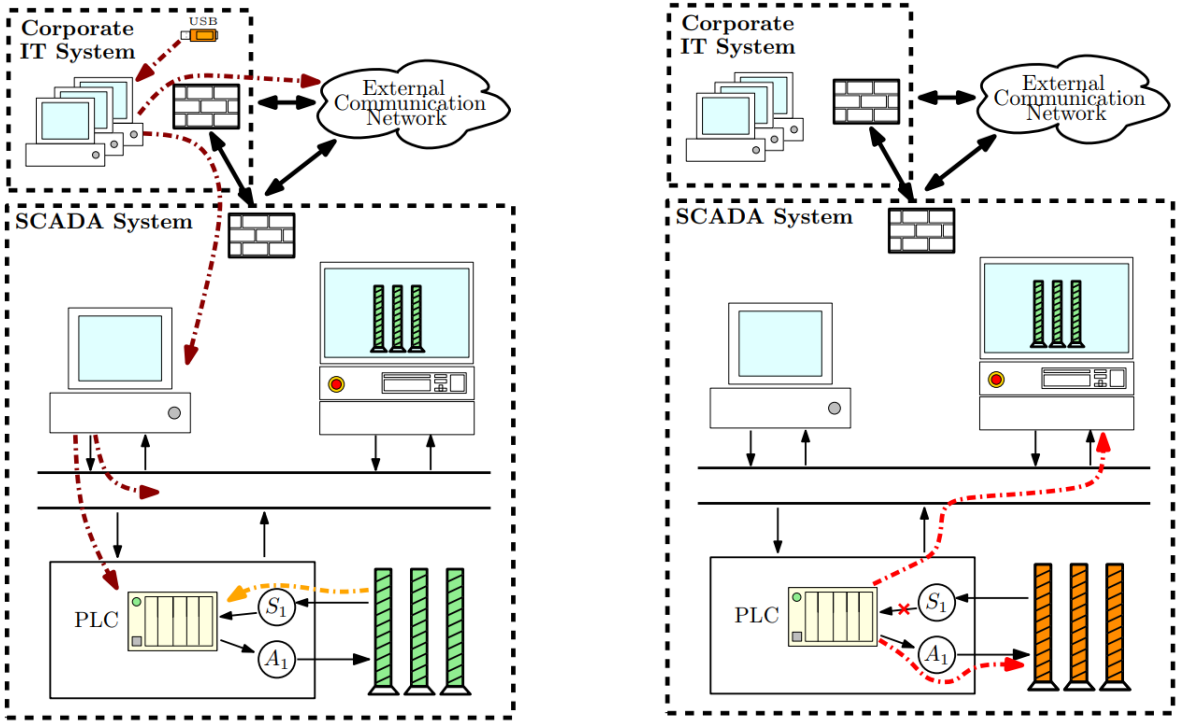
Targets: Windows, Siemens Step 7, and PLCs connected to variable-frequency drives

Exploited **4 zero-day flaws**

Goal: Harm centrifuges at uranium enrichment facility in Iran

[Kushner, “The Real Story of Stuxnet”, IEEE Spectrum, 2013]

Example 1: The Stuxnet Worm (2010)



(a) Infection and data recording.

(b) Covert sabotage.

[Teixeira, “Towards cyber-secure and resilient networked control systems,” PhD thesis, KTH, 2014]

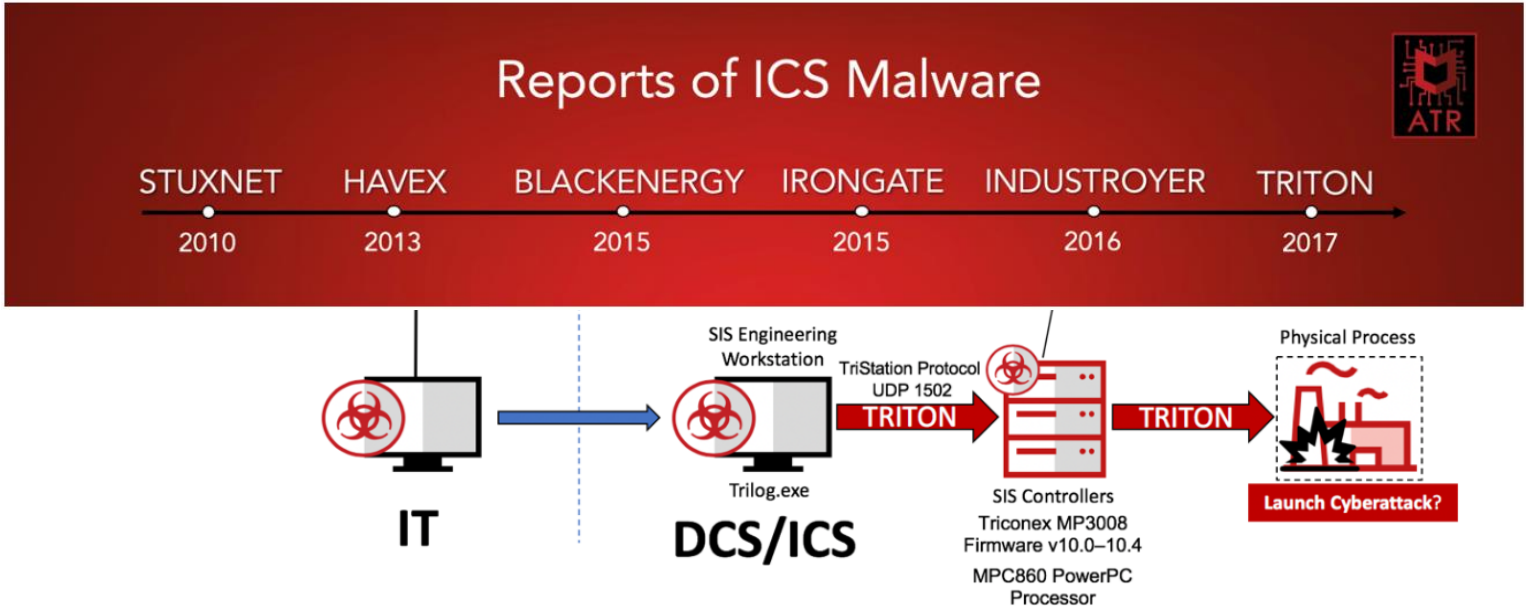
[https://www.mcafee.com/blogs/other-blogs/mcafee-labs/triton-malware-spearheads-latest-generation-of-attacks-on-industrial-systems/]

Example 2: Triton Malware (2017)

Triton framework

Triton targeted the Triconex safety controller, distributed by Schneider Electric. Triconex safety controllers are used in 18,000 plants (nuclear, oil and gas refineries, chemical plants, etc.), according to the company. Attacks on SIS require a high level of process comprehension (by analyzing acquired documents, diagrams, device configurations, and network traffic). SIS are the last protection against a physical incident.

The attackers gained access to the network probably via spear phishing, according to an investigation. After the initial infection, the attackers moved onto the main network to reach the ICS network and target SIS controllers.



Example 3: Events in Ukraine (2015) and USA (2019)

Analysis confirms coordinated hack attack caused Ukrainian power outage

BlackEnergy was key ingredient used to cause power outage to at least 80k customers.

by **Dan Goodin** - Jan 11, 2016 5:42am GMT

Share Tweet Email 33

The people who carried out last month's first known hacker-caused power outage used highly destructive malware to gain a foothold into multiple regional distribution power companies in Ukraine and delay restoration efforts once electricity had been shut off, a newly published analysis confirms.

The malware, known as BlackEnergy, allowed the attackers to gain a foothold on the power company systems, said the report, which was published by a member of the SANS

FURTHER READING

June 18, 2019

Hacking the Russian Power Grid

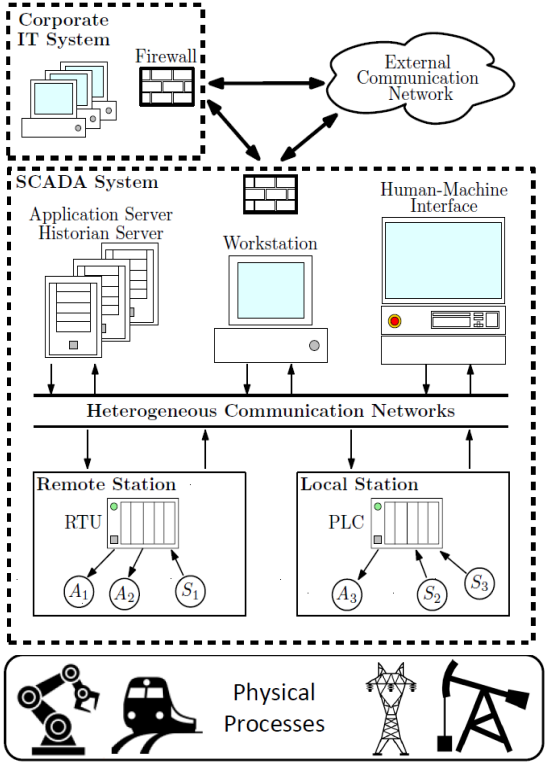
Attacks by the United States risk escalating a digital Cold War and renew questions about whether certain targets should be off limits in cyber conflict.

Hosted by Michael Barbaro; produced by Eric Krupke and Luke Vander Ploeg, with help from Jessica Cheung; and edited by Larissa Anderson

Typical ICS Security Vulnerabilities

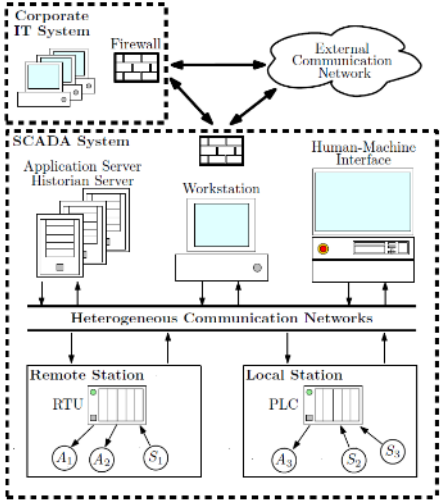
- Computers do not have adequate protection
  - No anti-virus or intrusion detection, USB-ports accessible
- Communication links lack basic security features
  - No encryption, no authentication
- Lack of physical protection
  - PLCs and RTUs accessible
  - Zero-day flaws

Systems designed for an "old" threat landscape!



[Cardenas *et al*., "Research challenges for the security of control systems", HotSec, 2008]

Security Challenges in ICS



Differences to traditional IT systems:

– **Patching and frequent updates are not well suited for control systems**

– **Real-time availability** (Strict operational environment)

– **Legacy systems** (Often no authentication, no encryption)

– **Protection of information and physical world** (Estimation and control algorithms)

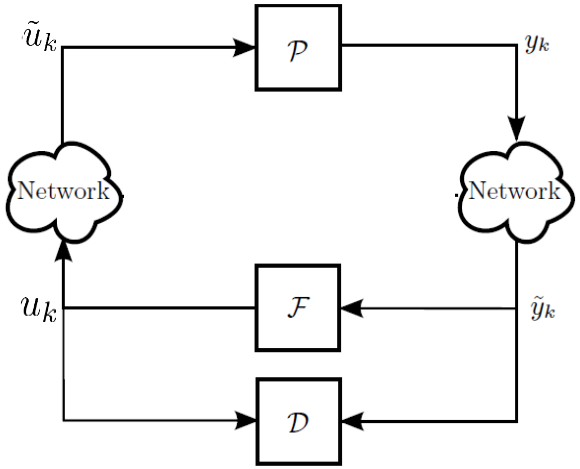
\+ **Simpler network dynamics** (Fixed topology, regular communication, limited number of protocols,…)

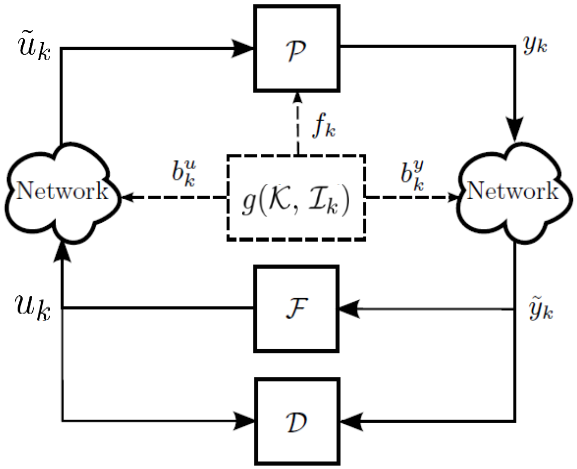
[Cardenas *et al*., “Research challenges for the security of control systems”, HotSec, 2008]

Outline

- Security challenges for industrial control systems
- **CPS security in networked control**
- Work in our group
  - Energy management systems, stealthy attacks, and security metrics
  - Confidentiality in the closed loop
- Summary and outlook

Networked Control System under Attack



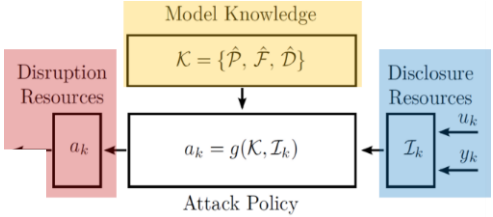


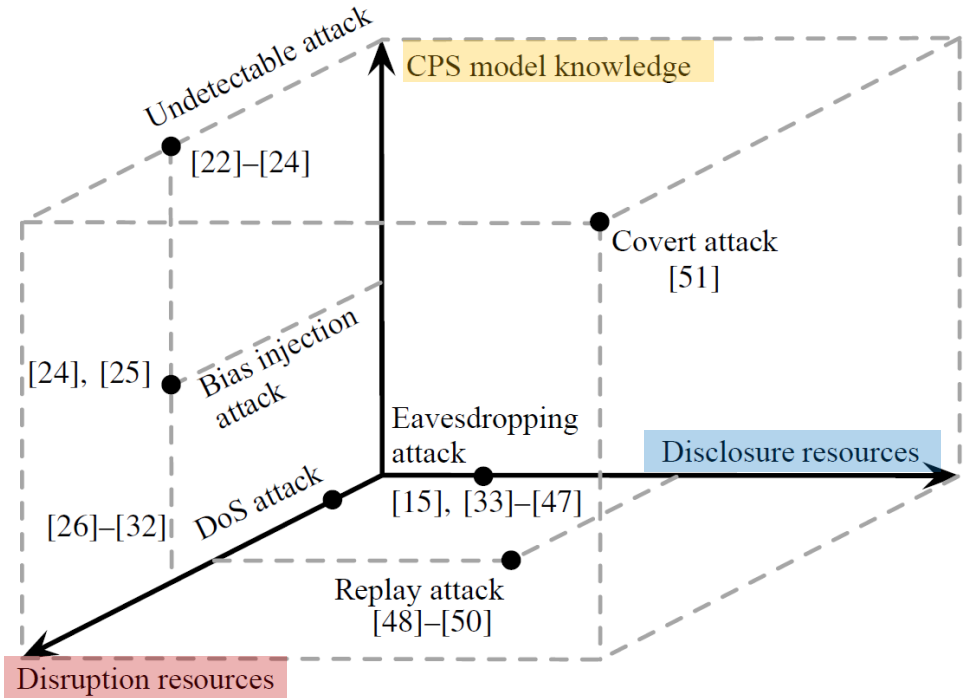
- Physical plant ($\mathcal{P}$)
- Feedback controller ($\mathcal{F}$)
- Anomaly detector ($\mathcal{D}$)

- Disclosure Attacks
- Physical Attacks f_k
- Data Injection Attacks b_k^y, b_k^u

[Teixeira *et al*., “A secure control framework for resource-limited adversaries”, Automatica, 2015]

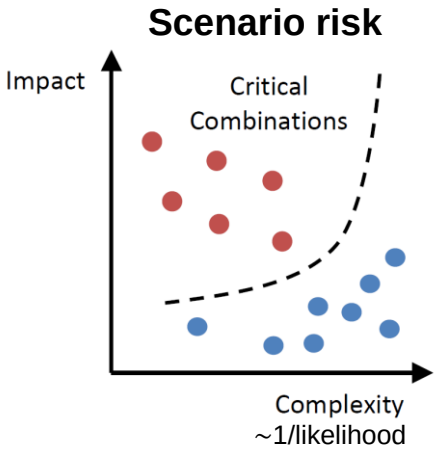
CPS Attack Space





[Chong, Teixeira, Sandberg, “A Tutorial Introduction to Security and Privacy for Cyber-Physical Systems,” ECC, 2019]

Tools for Risk Mitigation



- **Prevention** (decrease likelihood by reducing vulnerability)
  - Watermarking and Moving Target Defense
  - Coding and Encryption Strategies
  - *Rational Security Allocation*
  - *Confidentiality Protection by Noise Injection*
- **Detection** (continuous anomaly monitoring)
  - *Tuning of Detector Thresholds*
  - Secure State Estimation
  - *Watermarking and Moving Target Defense*
  - Robust Statistics
- **Treatment** (compensate for, or neutralize, detected attack)
  - *Secure State Estimation*
  - Countering DoS Attacks
  - Robust Statistics
  - Controller update

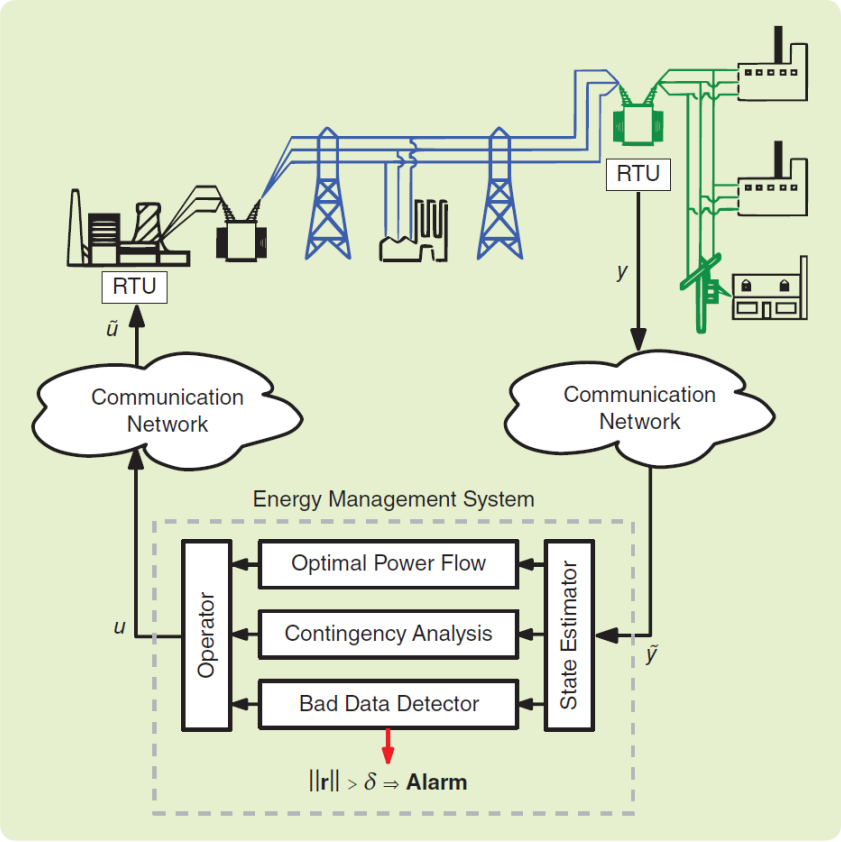
[Chong, Teixeira, Sandberg, “A Tutorial Introduction to Security and Privacy for Cyber-Physical Systems,” ECC, 2019]

Work in our Group

1. Energy management systems, stealthy attacks, and security metrics

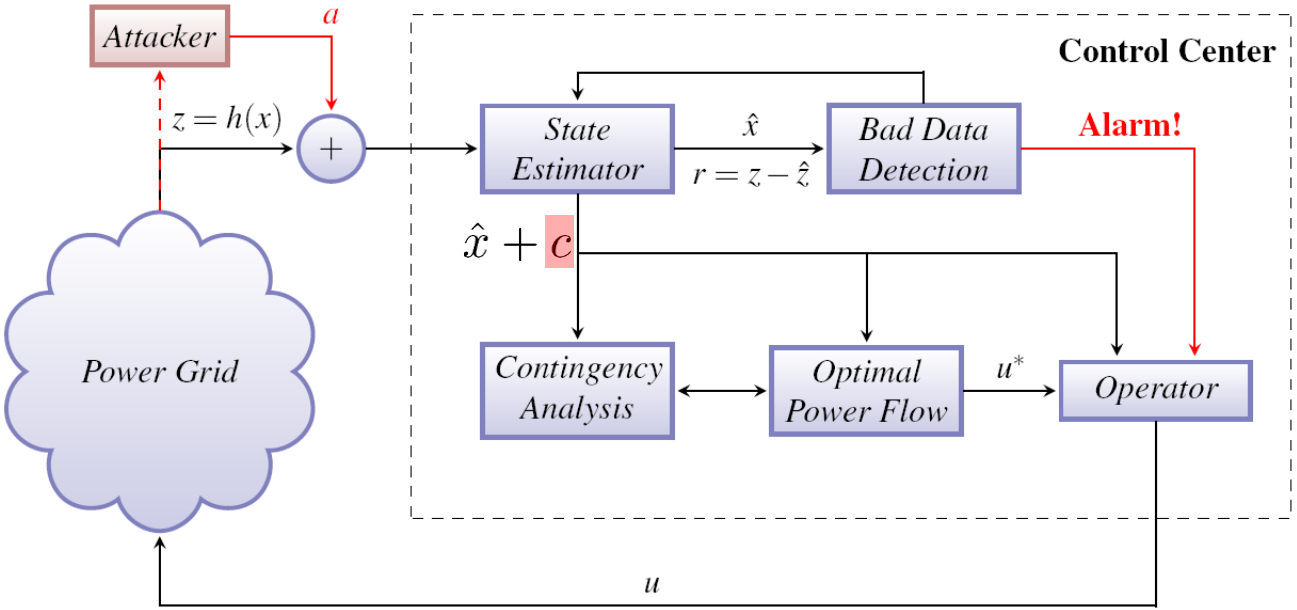
2. Confidentiality in the closed loop

Transmission Power System State Estimator



[Teixeira *et al.*, “Secure control systems: A quantitative risk management approach”, IEEE CSM, 2015]

Attacker Model and Bad Data Detection in Control Center



- **Scenario:** Attacker injects **malicious data** a to induce bias c in state estimate
- Typically $\dim(\text{measurement}) \gg \dim(\text{state})$. *Does attacker have to corrupt all sensors to remain undetected?*

Security Index

- Adversary launches undetectable attack against sensor channel i

$$\alpha_i := \min_c \|a\|_0 \quad \text{(sparsest possible attack)}$$
$$\text{s.t.} \quad a = Hc \quad \text{(undetectable)}$$
$$a_i = 1 \quad \text{(targets sensor } i\text{)}$$
$$\|a\|_0 := |\{a_k;\ a_k \neq 0\}|$$

[Sandberg *et al.*, "On Security Indices…", SCS, 2010]

- Quantifies complexity of "least-effort undetectable attack" on sensor i.
- **Example:** $\alpha_1 = 2$ undetectable attack against sensor 1 involves *at least two* sensors in total
- Efficient min-cut/max-flow algorithm for computation exists

[Hendrickx *et al.*, "Efficient computations of a security index…", IEEE TAC, 2013]

Undetectable and Identifiable (Correctable) Sensor Attacks

Theorem: Suppose that the attacker can manipulate at most q sensors simultaneously ($\|a\|_0 \leq q$).

i. There exists undetectable attacks against sensor $i \Leftrightarrow \; q \geq \alpha_i$

ii. All attacks are i-identifiable (sensor i correctable) $\Leftrightarrow q \; < \alpha_i/2$

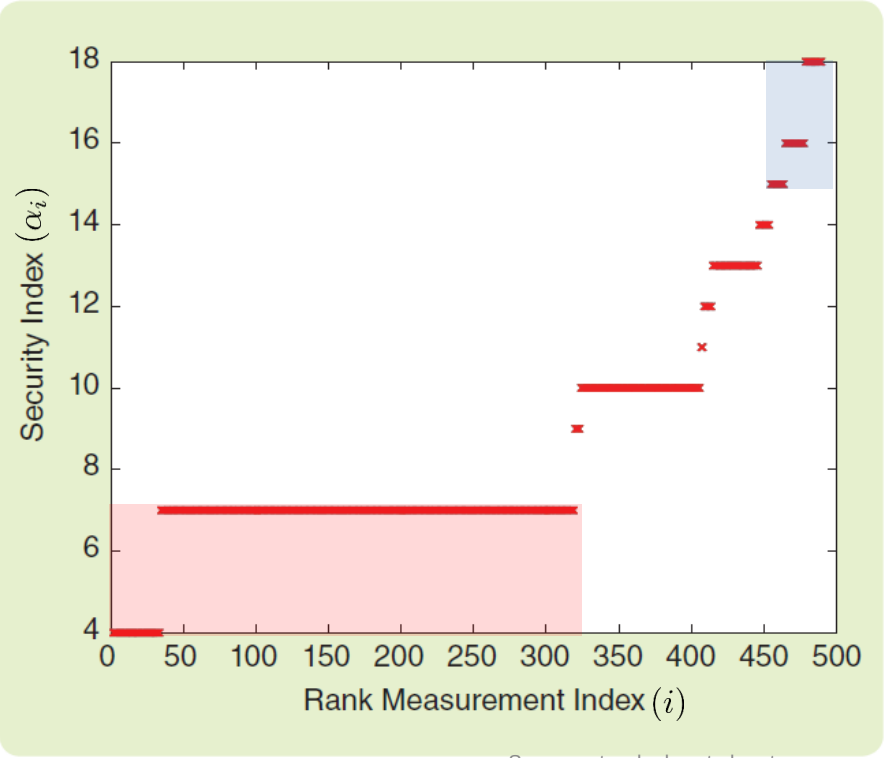
iii. All attacks are identifiable (all sensors correctable) $\Leftrightarrow q \; < \min_i \alpha_i/2$

- Proofs based on compressed-sensing and coding-theory type arguments

[Sandberg and Teixeira, “From Control System Security Indices to Attack Identifiability”, SoSCYPS, 2016]

Example: Power System State Estimator for IEEE 118-bus System

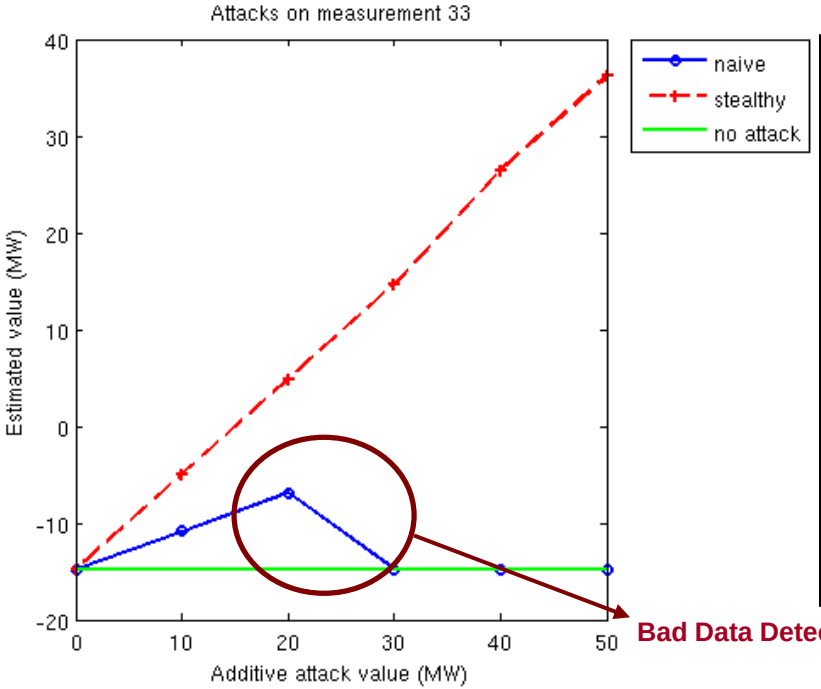
Suppose number of attacked elements is $q \leq 7$. Theorem yields:



- Sensors susceptible to undetectable attacks
- Sensors where attacks are correctable
- Other sensor attacks are (in principle) detectable, but not correctable

[Teixeira *et al.*, “Secure control systems: A quantitative risk management approach”, IEEE CSM, 2015]

Experiments on SCADA/EMS Testbed



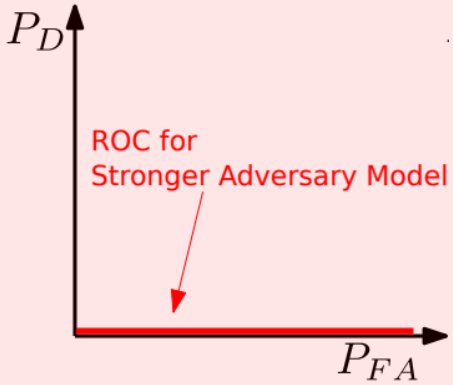
Bad Data Detected & Removed

False value (MW)	Estimated value (MW)	# BDD Alarms
-14.8	-14.8	0
35.2	36.2	0
85.2	86.7	0
135.2	137.5	0
185.2	Non convergent	-

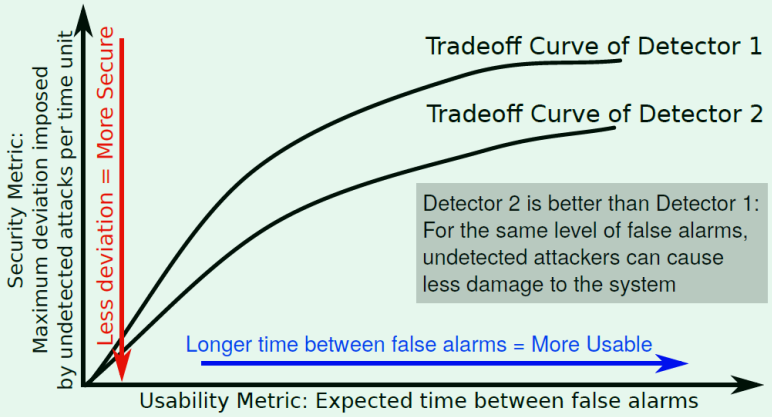
- Attacks of 150 MW (≈55% of nominal value) pass undetected in a real system!

[Teixeira, *et al*., “A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator”, IFAC WC, 2011]

New Detector Tuning Tools – No ROC!



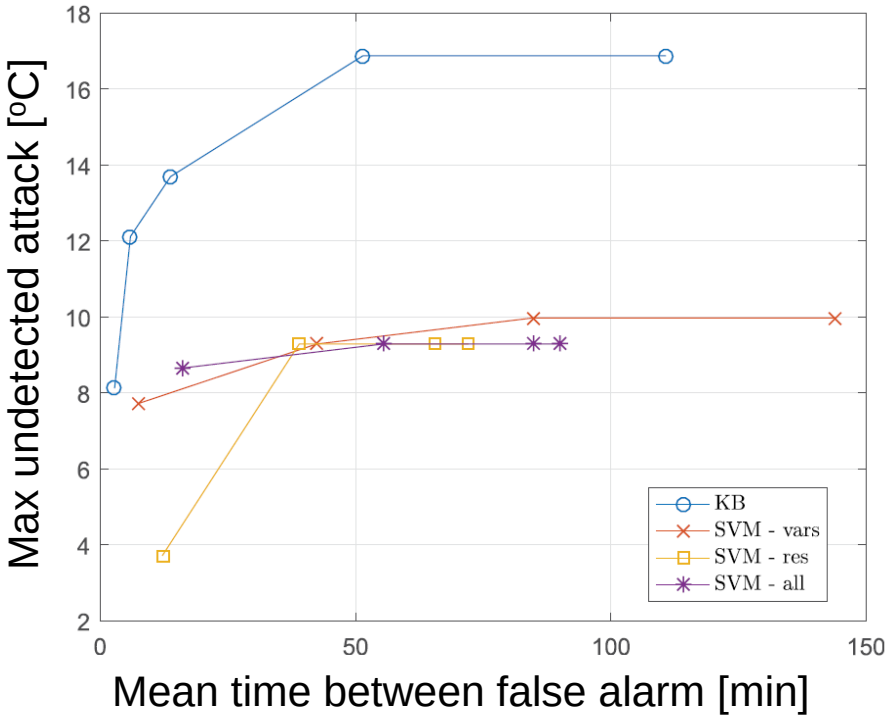
- Traditional ROC (Receiver Operating Characteristic) curve unsuitable for tuning detectors in this threat landscape



- New proposed metric for detector tuning. Requires tools for **impact estimation of stealthy attacks**

[Urbina, *et al*., “Limiting the Impact of Stealthy Attacks on Industrial Control Systems”, ACM CCS’16, 2016]
[Umsonst, “Stealthy Sensor Attacks on Feedback Systems: Attack Analysis and Detector-based Defense Mechanisms”, PhD thesis, KTH, 2021]

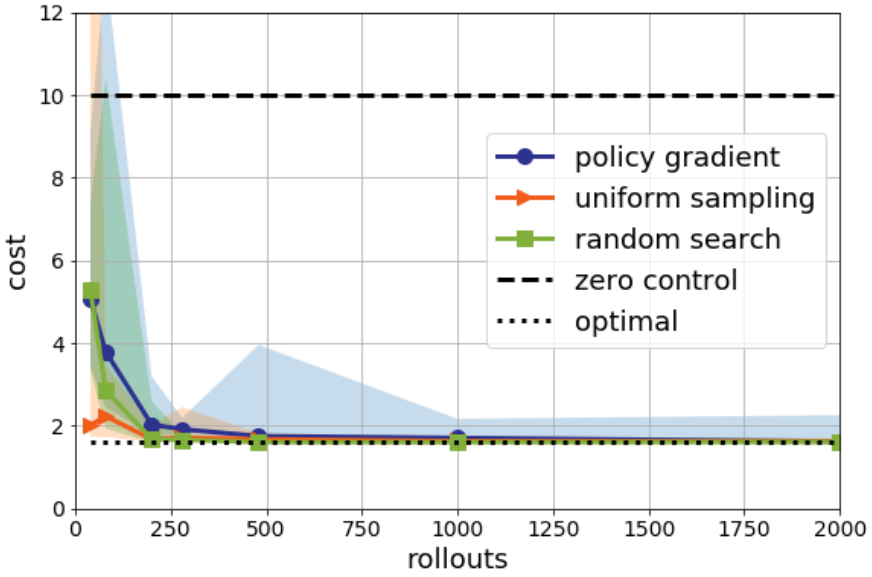
Attack Detector Comparison: NIMBUS Microgrid



- Data-based (1-SVM) detectors restricts attacker more
- Model-based (KB) detector only checks “physicality” of time series
- Data-based detector also checks for unusual operation

[Paridari, *et al.*, “A Framework for Attack-resilient Industrial Control Systems”, Proc. IEEE, 2018]

Safe, Secure, and Model-free Control?



Performance of model-free reinforcement learning applied to linear dynamical control problem

[B. Recht, “Updates on Policy Gradients”, arg min blog, 2018]

- Traditional system identification (10 data samples) + linear-quadratic control **converged to optimal after 1 rollout**. Exploiting known model structure is powerful
- **Data samples are not “free” in many CPS applications!** Comes with safety risk and costs

Work in our Group

1. Energy management systems, stealthy attacks, and security metrics

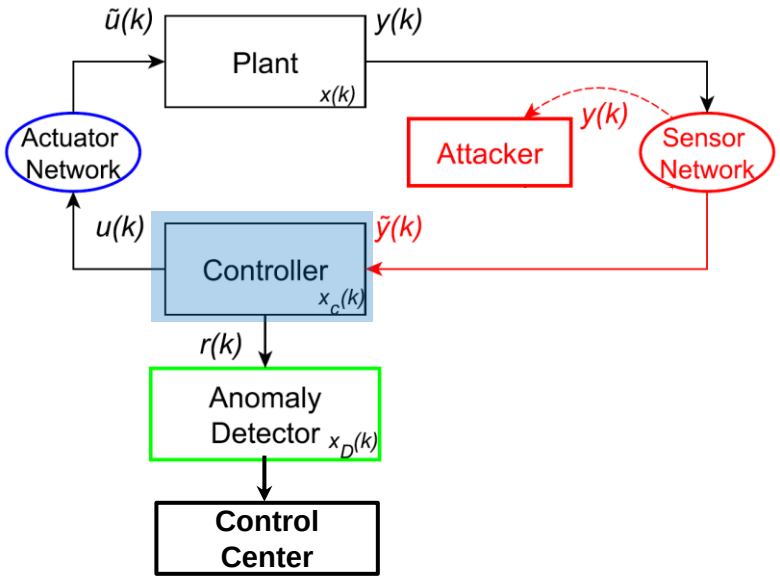
2. Confidentiality in the closed loop

Confidentiality of Controller States

- **Scenario:** Attacker has
  - Access to models of Plant and Controller (but **not** access to their internal states), and;
  - Access to sensor network

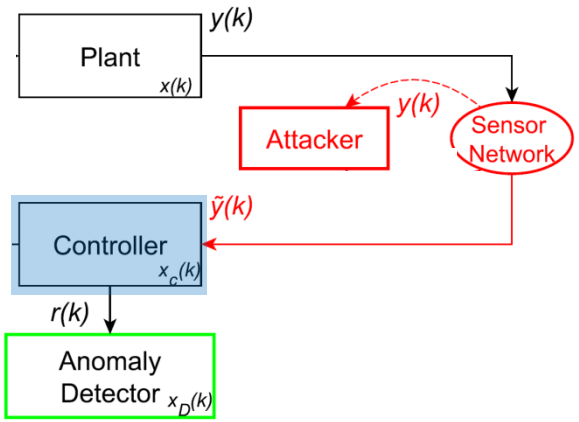
- **Problem:** Can the Attacker exactly track the internal state x_c of the controller?

- **Theorem:** 'No' if, and only, if the controller is (open-loop) unstable!

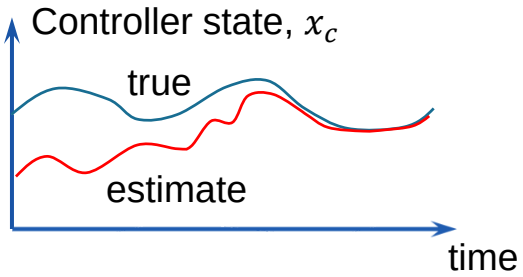


[Umsonst, Sandberg, "On the confidentiality of controller states under sensor attacks", Automatica, 2021]

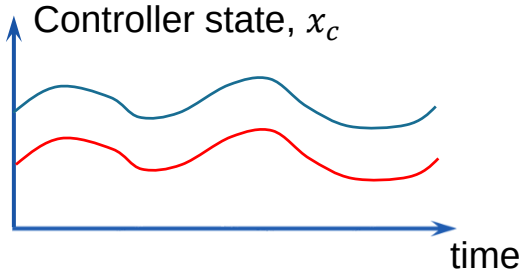
The (Simple) Open-Loop Case



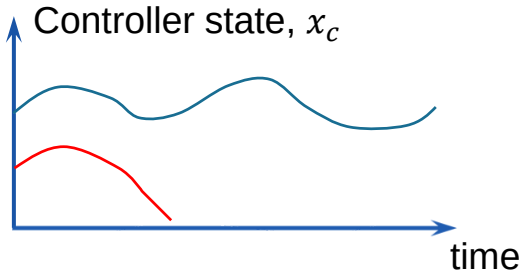
Controller stable:



Controller marginally stable:

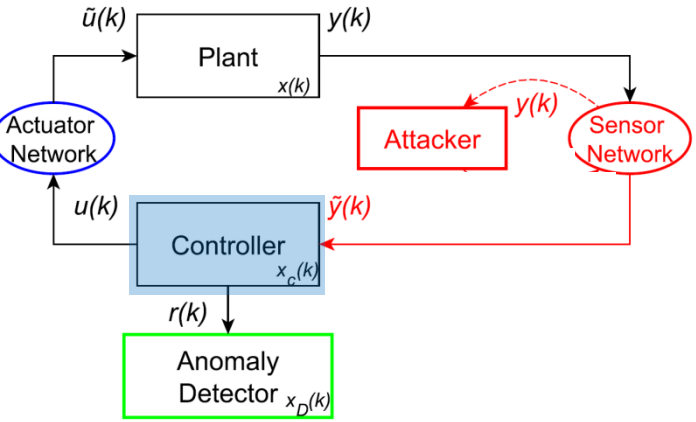


Controller unstable:

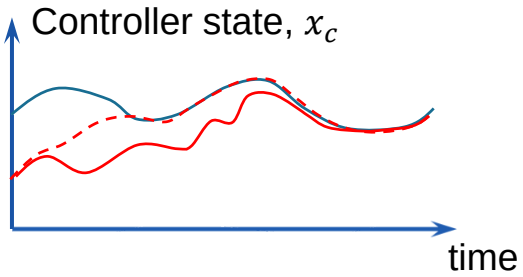


The Closed-Loop Case

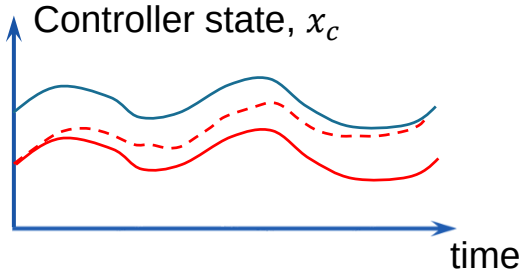
Controller *signals* side information through physical plant!



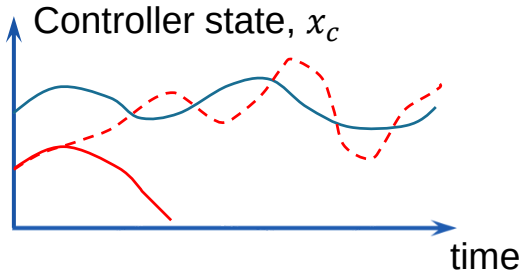
Controller stable:



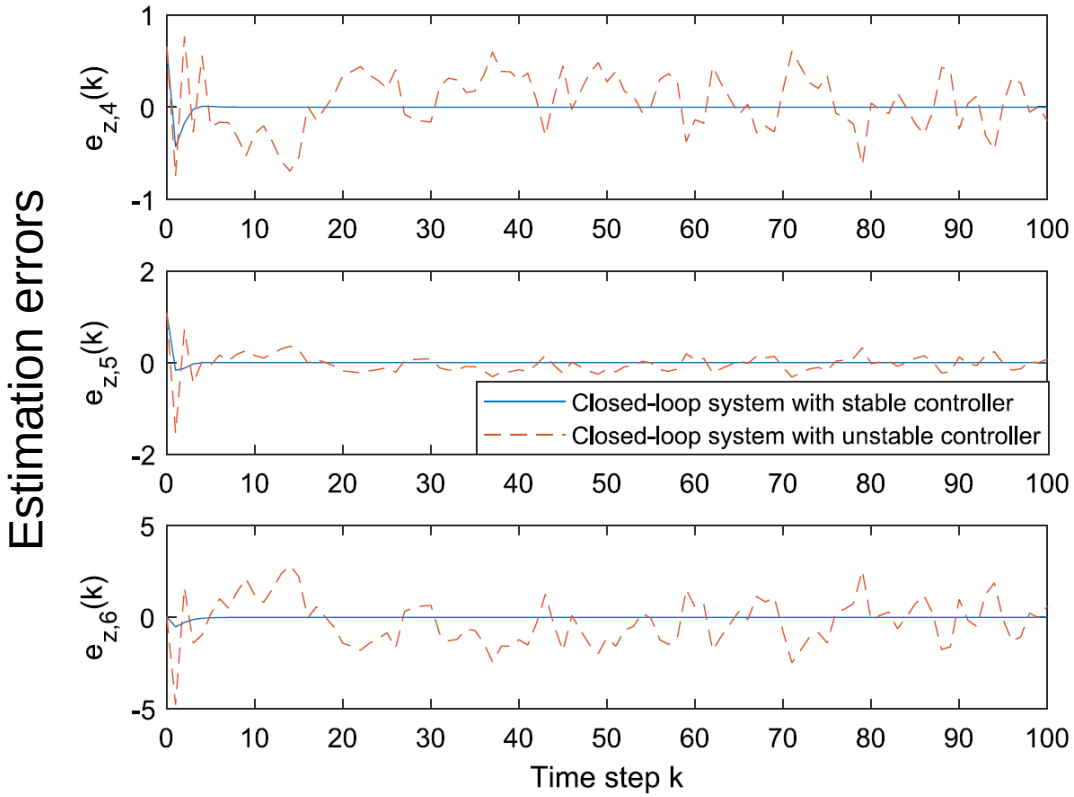
Controller marginally stable:



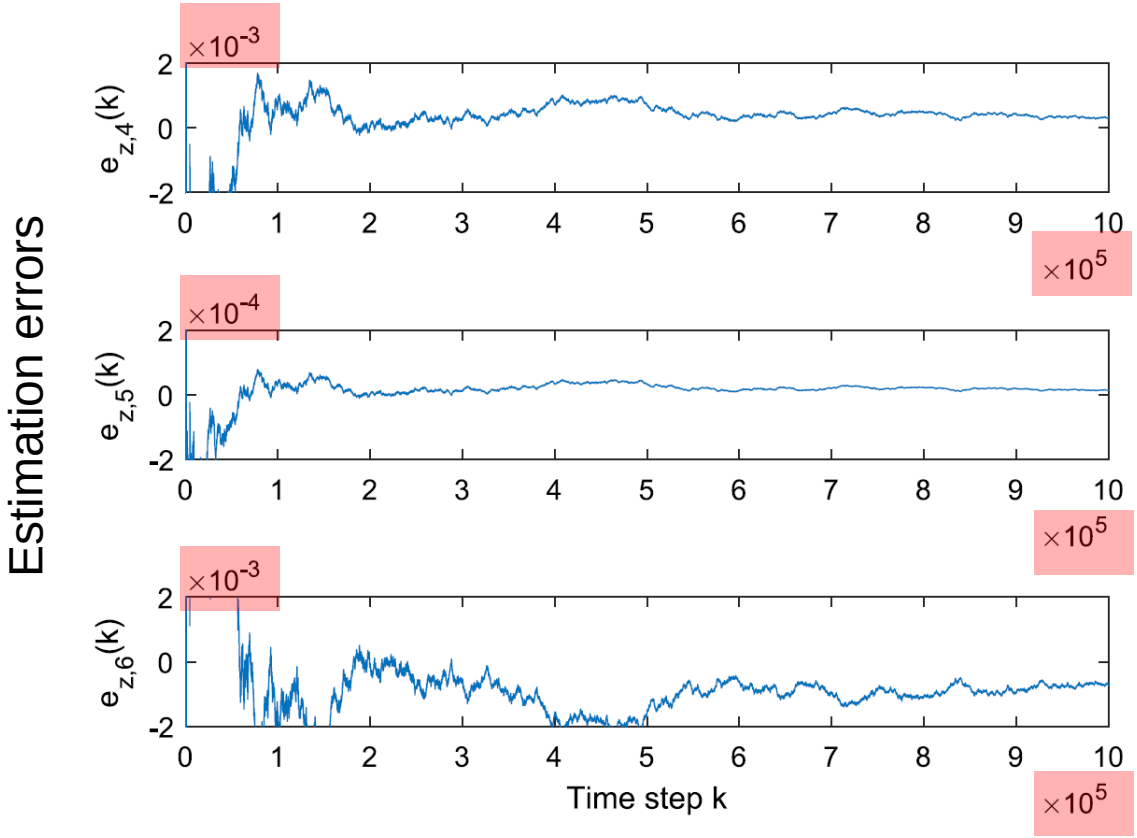
Controller unstable:



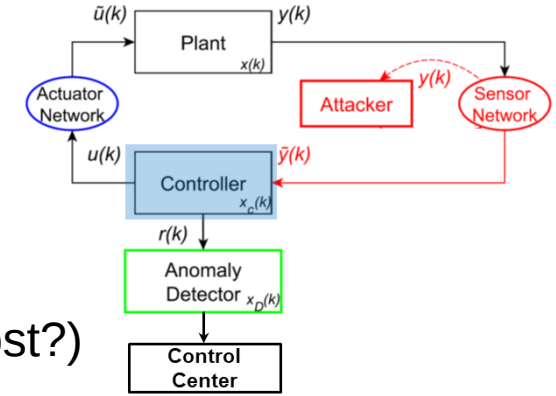
Examples with Noise



Example with Noise (Marginally Stable Case)



Discussion

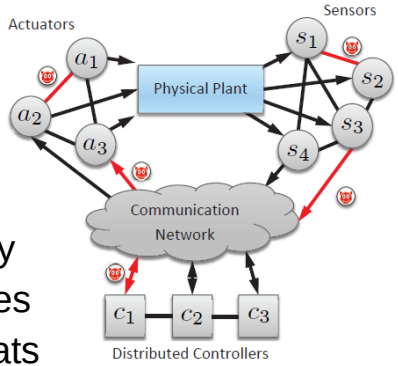


- Marginally stable case is not uncommon: Many (most?) controllers are P**I** (proportional-**integral**)!

- How to ensure confidentiality of controller?
  - Encrypt sensors…
  - Use unstable controller (dangerous and performance suffers)
  - Noise injection at controller input (performance suffers)

- If Attacker has access also to control signal u, attacker can always estimate controller state x_c exponentially fast

Final Takeaways



- Security increasingly important in control and monitoring systems
  - Dramatic attacks reported in media: nuclear, water, gas, process industry
  - Critical infrastructures, legacy components, heterogeneous infrastructures
  - Dedicated malware since more than 10 years, advanced persistent threats
  - *IT security necessary, not sufficient. Think defense in depth!*

- Work in our group
  - Energy management systems, security metrics, detector tuning and design, resilient estimation and control
  - Confidentiality and privacy in the closed loop

- Topics for future work
  - Tools from Machine Learning and AI
  - Fundamental design trade-offs (performance – security – safety – privacy)
  - Correct architecture for safe and secure distributed, critical control?