Secure control systems: An overview

André Teixeira
Uppsala University, Sweden
andre.teixeira@angstrom.uu.se

www.andre-teixeira.eu

Plenary Talk, SysTol'19, Sep. 19

Outline

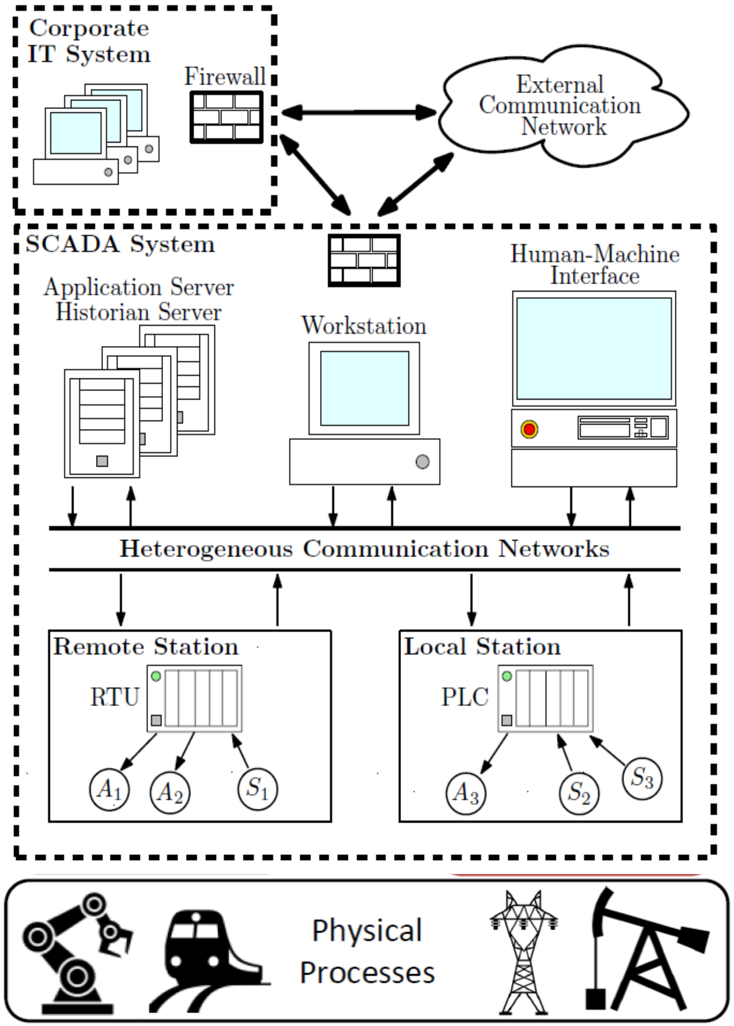
- Motivation

- Risk management
  - Scenario characterisation
  - Risk Analysis
  - Risk Mitigation

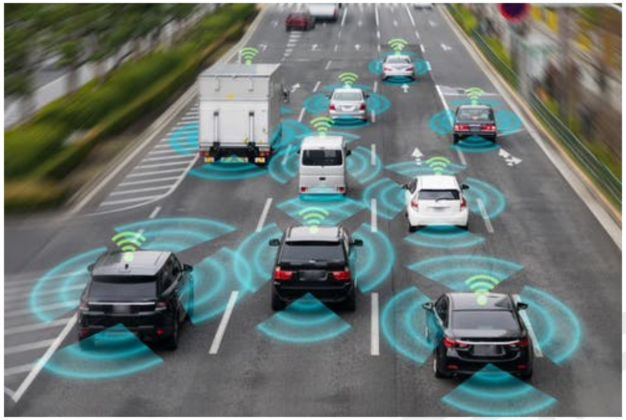
- Secure control: from analysis to design
  - Problem formulation
  - Metrics in Fault-Tolerant Control
  - Metrics in Secure Control
    - Computation and design problem

Cyber-Physical Systems

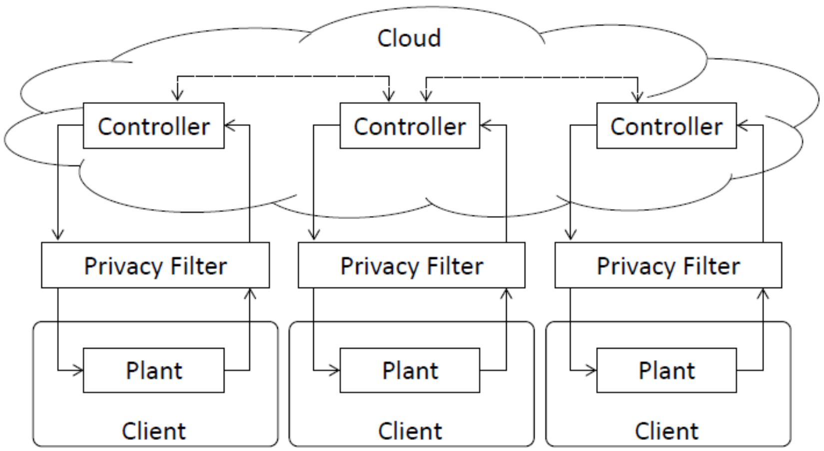
Industrial Control System (ICS)



Autonomous vehicles



Cloud-based Control and IoT

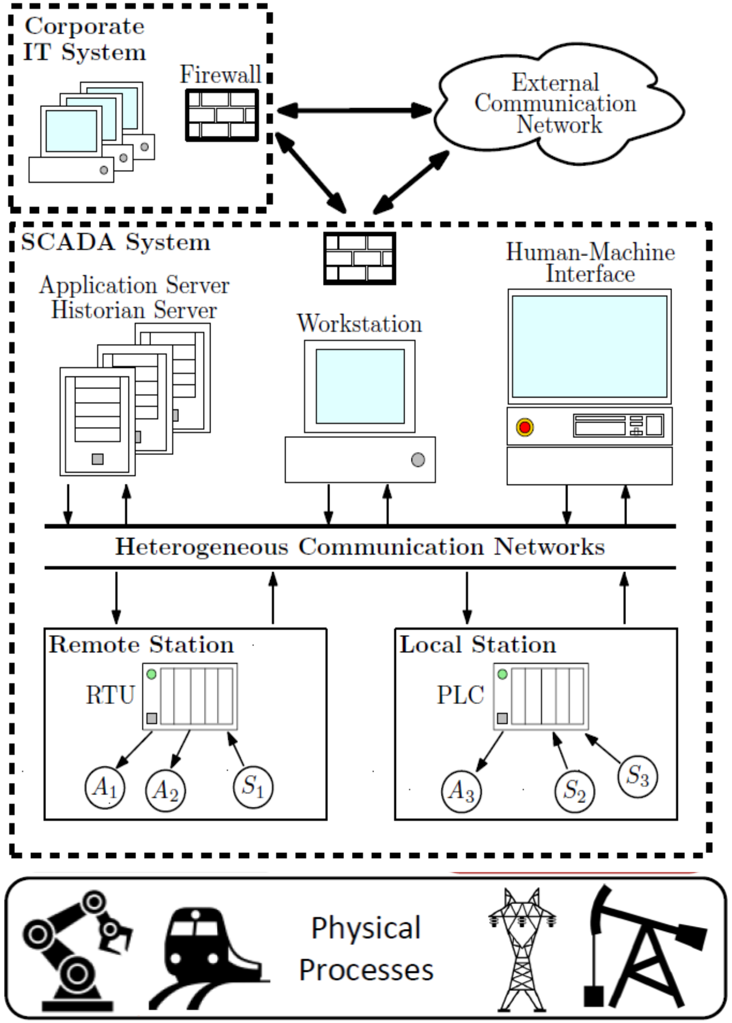


Typical ICS Vulnerabilities

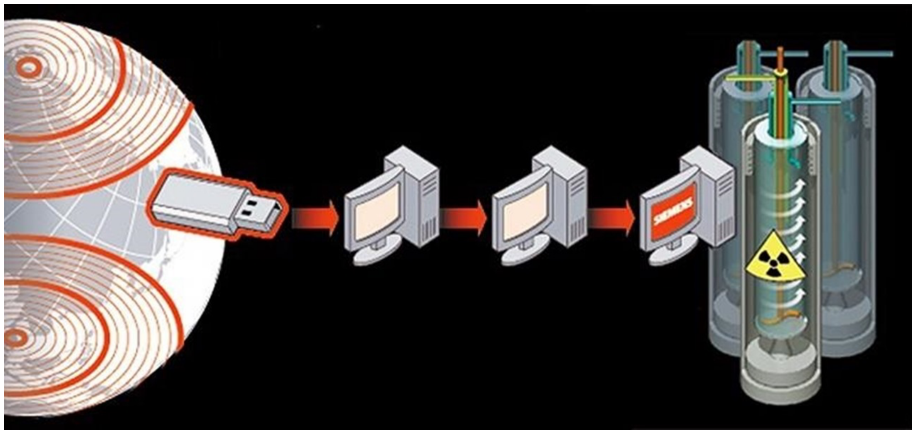
- Computers in control center do not have adequate protection
  - No anti-virus or intrusion detection, USB-ports accessible

- Communication links lack basic security features
  - No encryption or authentication

- Zero-day vulnerabilities



Example 1: Stuxnet (2010)



https://en.wikipedia.org/wiki/Zero_Days

Synopsis
Zero Days covers the phenomenon surrounding the Stuxnet computer virus and the development of the malware software known as "Olympic Games." It concludes with discussion over follow-up cyber plan Nitro Zeus and the Iran Nuclear Deal.

Zero Days

Theatrical release poster

Directed by	Alex Gibney
Written by	Alex Gibney
Distributed by	Magnolia Pictures
Release date	February 11, 2016 (Berlin) July 8, 2016 (US)
Running time	116 minutes
Country	United States
Language	English

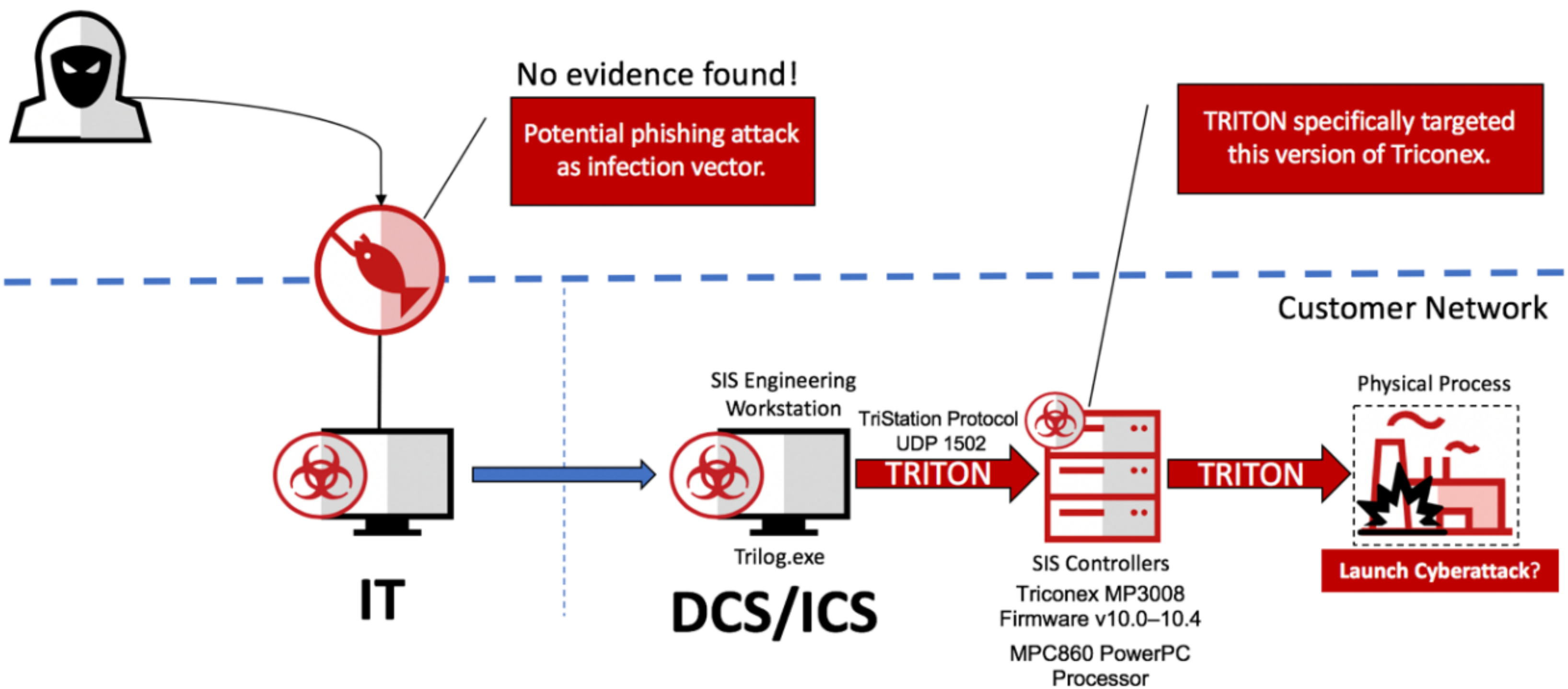
[https://securingtomorrow.mcafee.com/other-blogs/mcafee-labs/triton-malware-spearheads-latest-generation-of-attacks-on-industrial-systems/]

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.

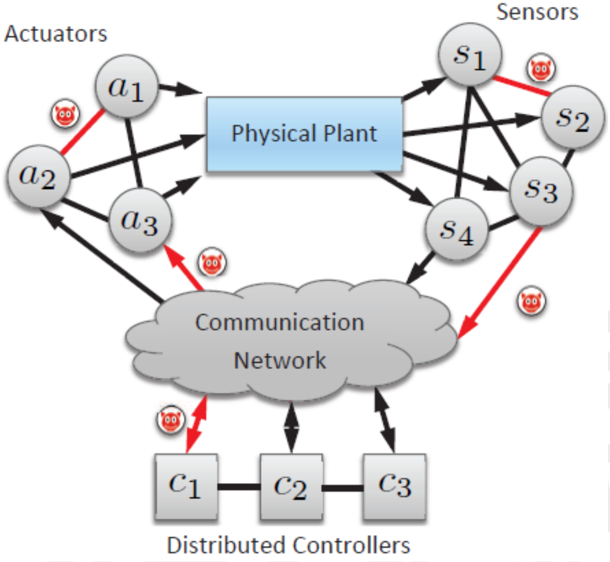
The attackers gained access to the network probably via spear phishing, according to an investigation. After the initial infection, the attackers moved onto the main network to reach the ICS network and target SIS controllers.



Cyber-Secure Control

Networked control systems

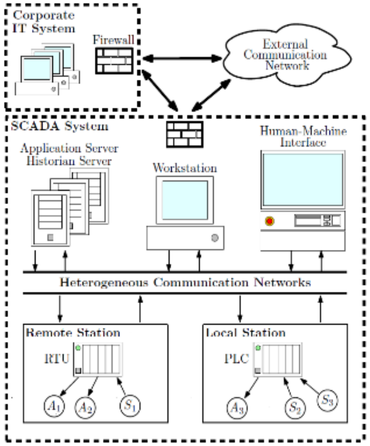
- are being **integrated with business/corporate networks**
- have many potential points of **cyber-physical attack**

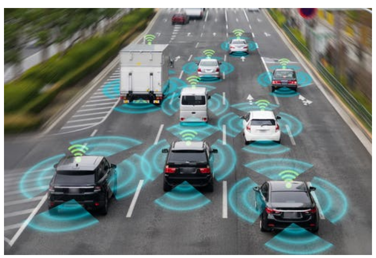


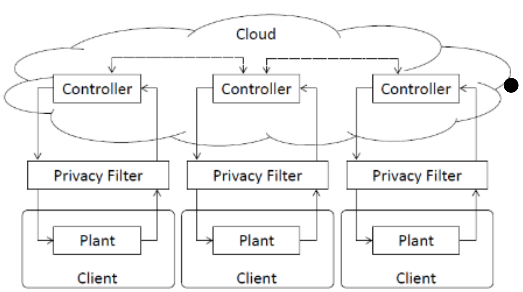
Need tools and strategies to understand and mitigate attacks:

- **Which threats** should we care about?
- **What impact** can we expect from attacks?
- **Which resources** should we **protect**, and how?

Is More Than IT Security and Fault Tolerance Needed?







- Clearly IT security and fault tolerance are needed: Authentication, encryption, firewalls, error correction, etc.

But not sufficient…

- **Interaction between physical and cyber systems** make control systems different from normal IT systems
- **Can we trust** the interfaces and channels are really secured? (see **OpenSSL** Heartbleed bug…)
- **Malicious actions can enter anywhere** in the closed loop and cause harm
- **Malicious attackers** have an **intent,** as opposed to faults, and can act strategically

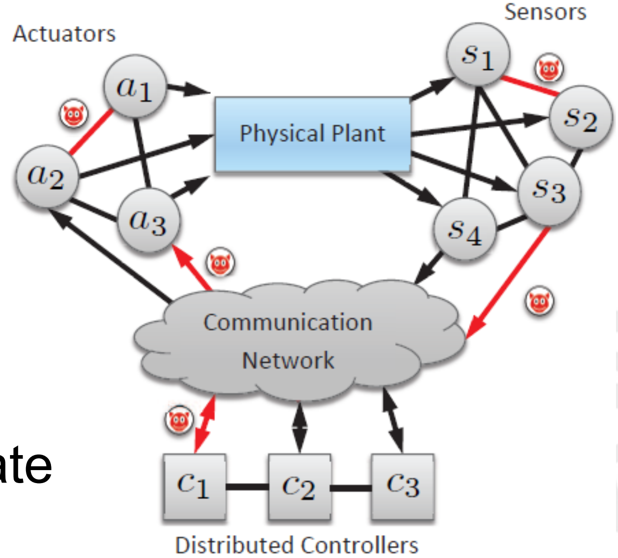
Outline

- Motivation

- **Risk management**
  - Scenario characterisation
  - Risk Analysis
  - Risk Mitigation

- Secure control: from analysis to design
  - Problem formulation
  - Metrics in Fault-Tolerant Control
  - Metrics in Secure Control
    - Computation and design problem

Cyber-Secure Control



Networked control systems

- are being **integrated with business/corporate networks**
- have many potential points of **cyber-physical attack**

Need tools and strategies to understand and mitigate attacks:

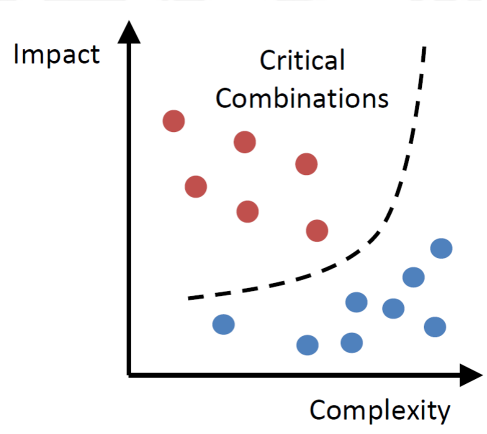
- **Which threats** should we care about?
- **What impact** can we expect from attacks?
- **Which resources** should we **protect**, and how?

- How to find answers**: Risk Management**

Defining Risk

Risk = (Scenario, Likelihood, Impact)

- **Scenario**
  – How to describe the system under attack?
- **Likelihood**
  – Interpretations:
    a) *Likelihood of attack in progress being successful (experts' assessment)*
    b) *Likelihood = 1*
    c) *~1/effort to conduct attack (or ~1 / complexity of attack)*
- **Impact**
  – What are the cyber-physical consequences of an attack?

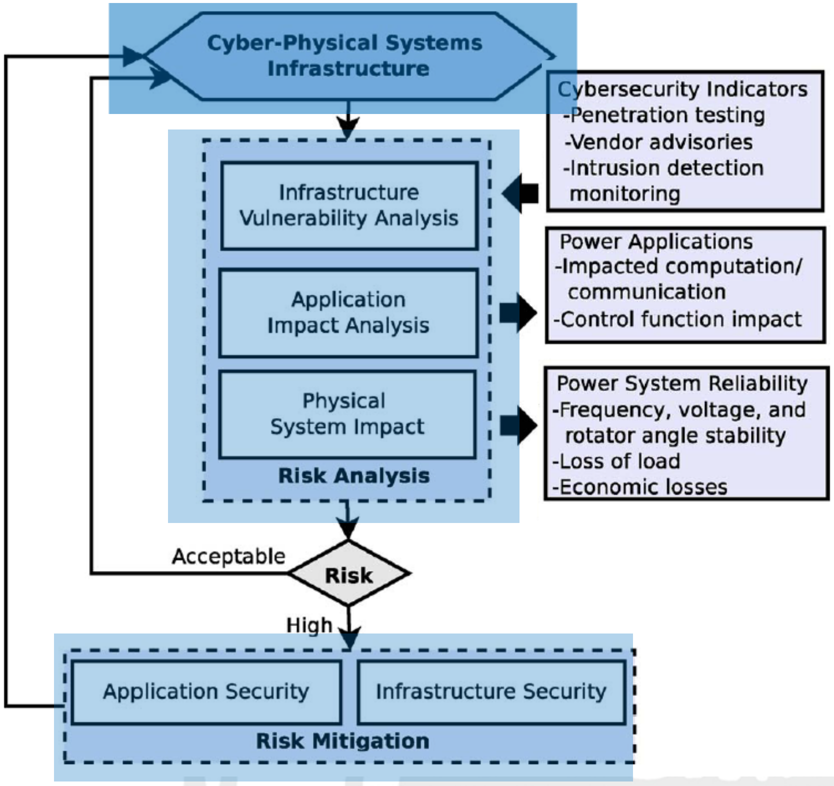


[Kaplan & Garrick, 1981], [Bishop, 2002]

Risk Management Cycle

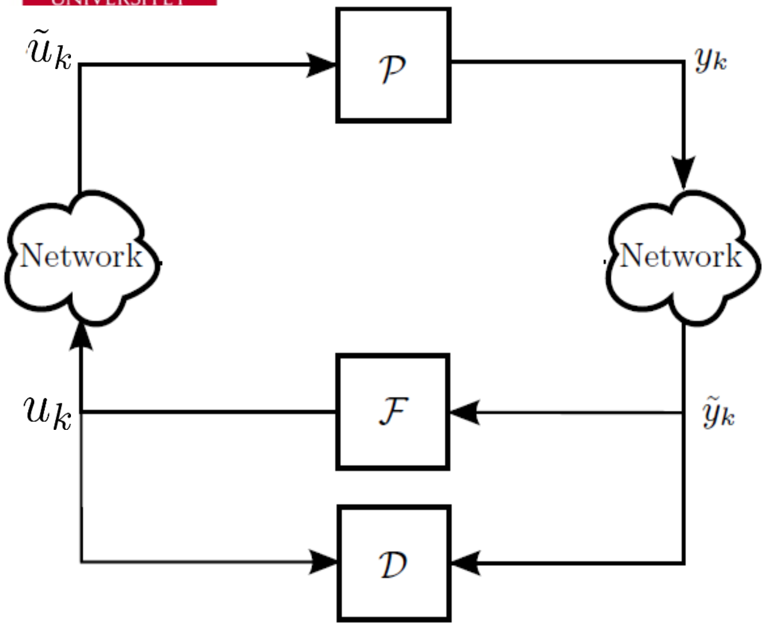
Main steps in risk management

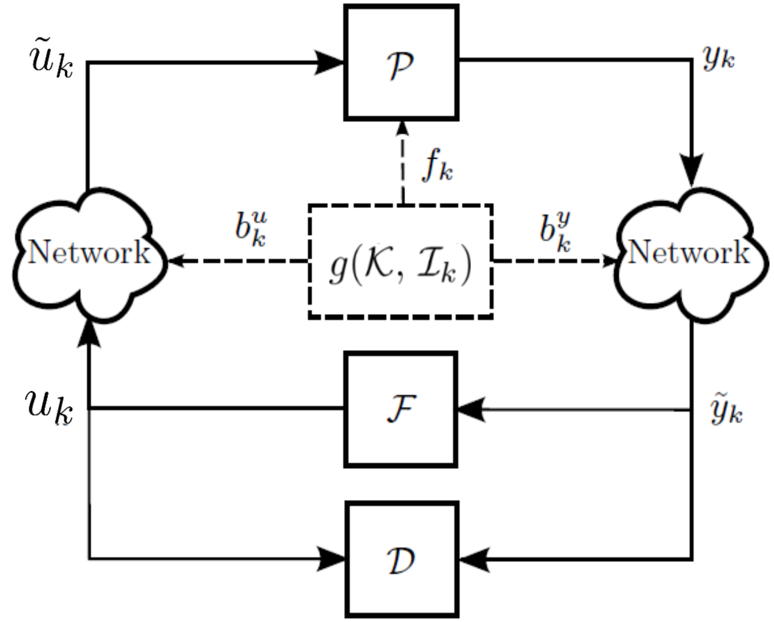
- Scenario characterization
  - Models, Scenarios, Objectives
- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment
- Risk Mitigation
  - Prevention, Detection, Treatment



[Sridhar *et al.*, Proc. IEEE, 2012]

Networked Control System under Attack





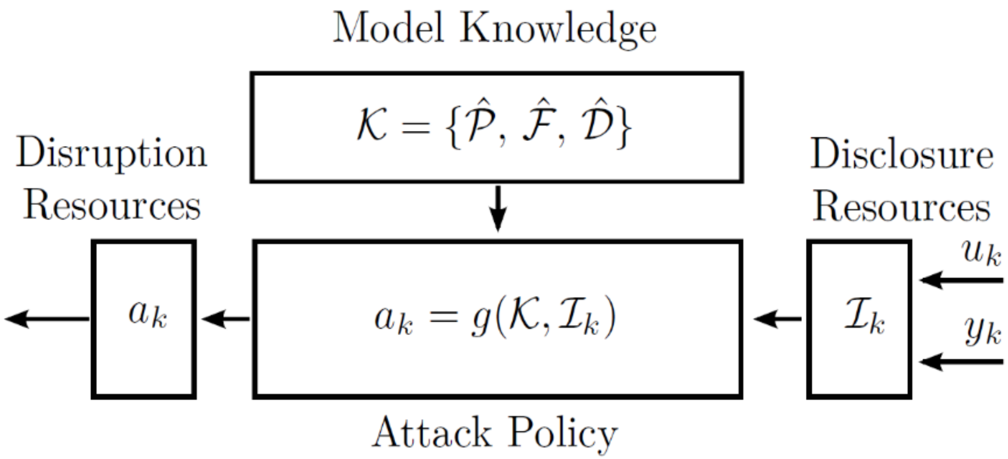
- Physical plant ($\mathscr{P}$)
- Feedback controller ($\mathscr{F}$)
- Anomaly detector ($\mathscr{D}$)
- Disclosure Attacks

- Physical Attacks f_k
- Data Injection Attacks

$$\tilde{u}_k = u_k + \Gamma^u b_k^u$$
$$\tilde{y}_k = y_k + \Gamma^y b_k^y$$

[Teixeira *et al.*, Automatica, 2015]

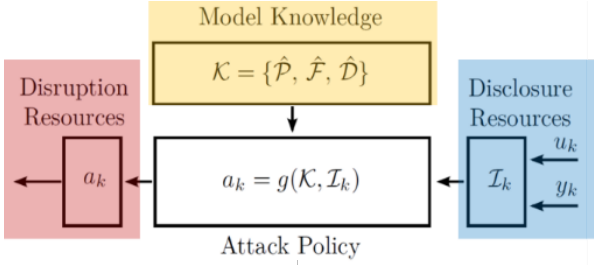
Adversary Model

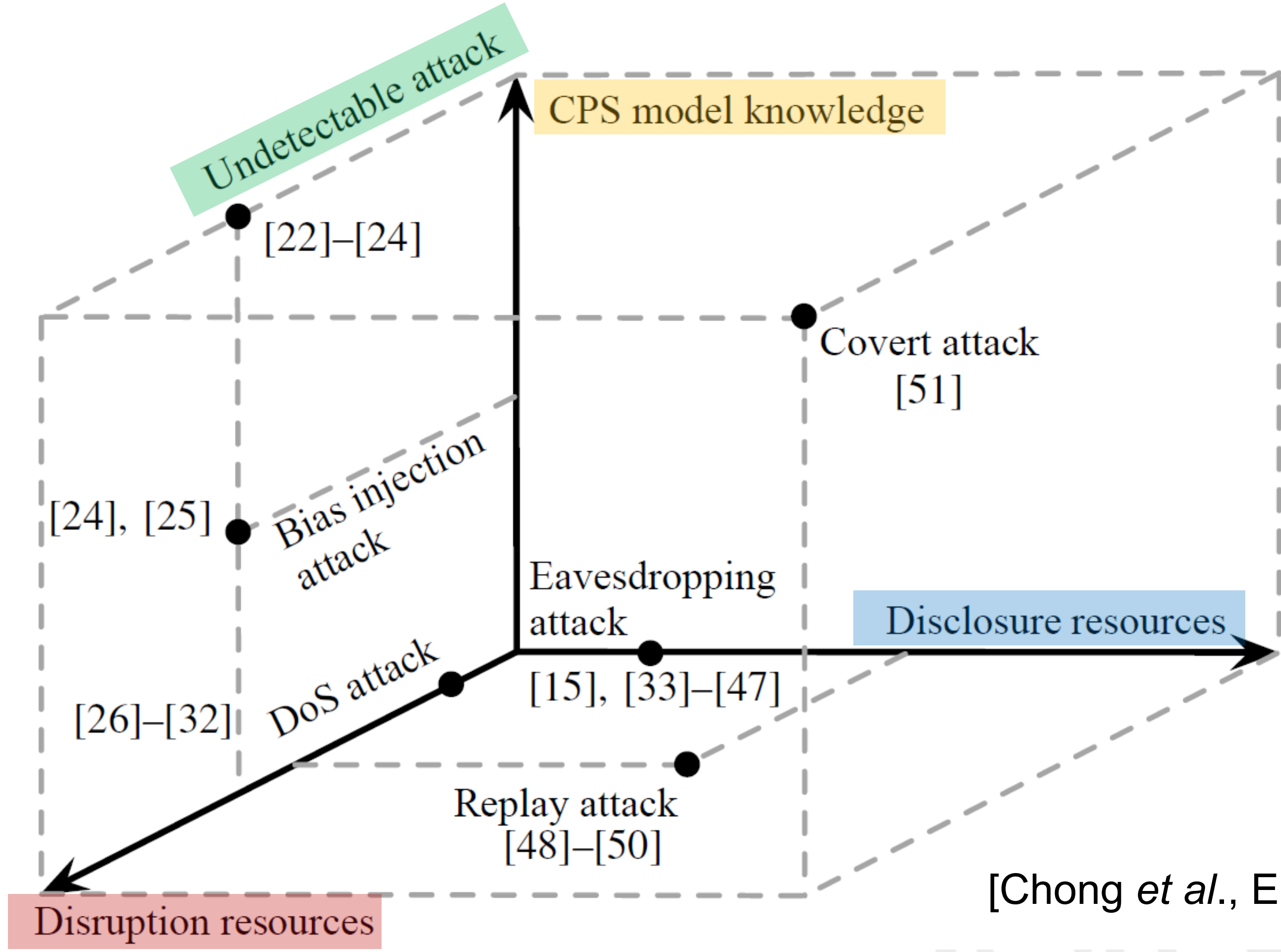


- **Attack policy:** Goal of the attack? Destroy equipment, increase costs, *remain undetected*…
- **CPS model knowledge:** Adversary knows models of plant and controller? Possibility for stealthy attacks…
- **Disruption/disclosure resources:** Which channels can the adversary access?

[Teixeira *et al*., Automatica, 2015]

Attack Space





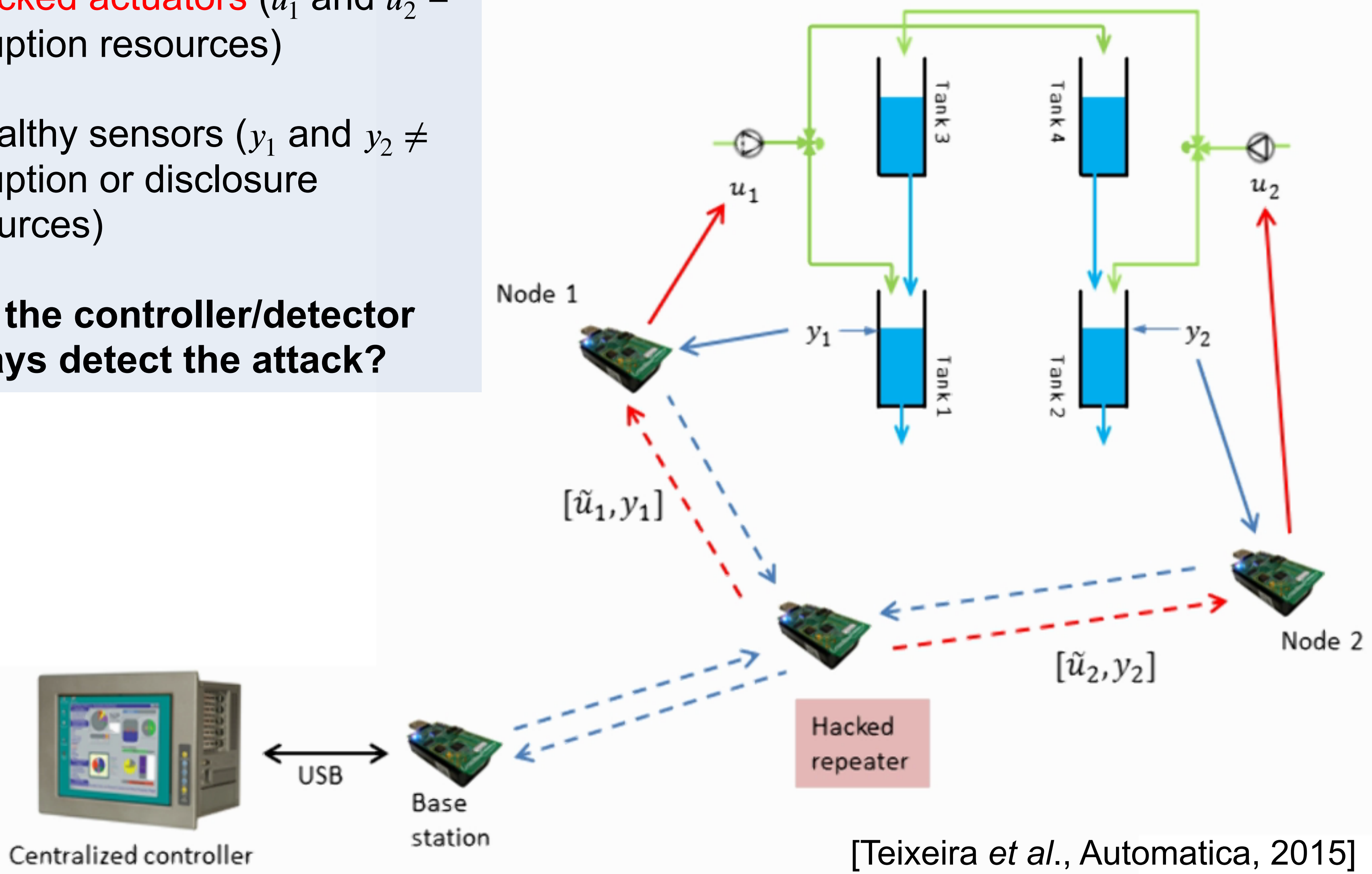
[Chong *et al.*, ECC, 2019]

Example: Undetectable Water

2 hacked actuators (u_1 and u_2 = disruption resources)

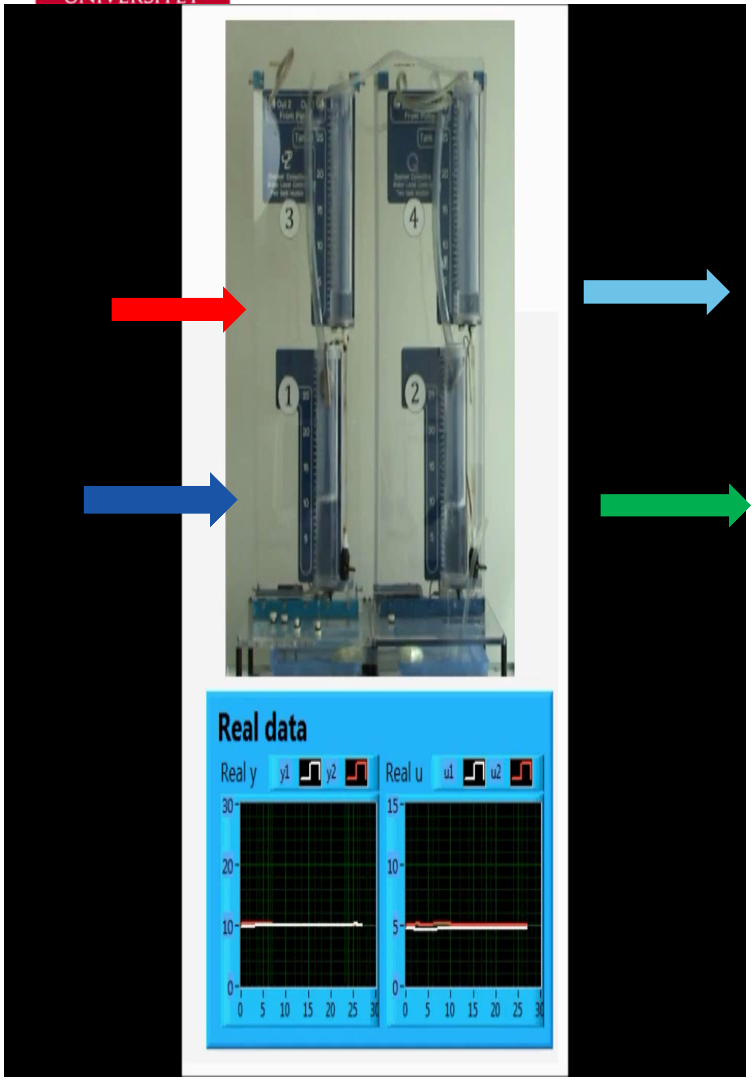
2 healthy sensors (y_1 and $y_2 \neq$ disruption or disclosure resources)

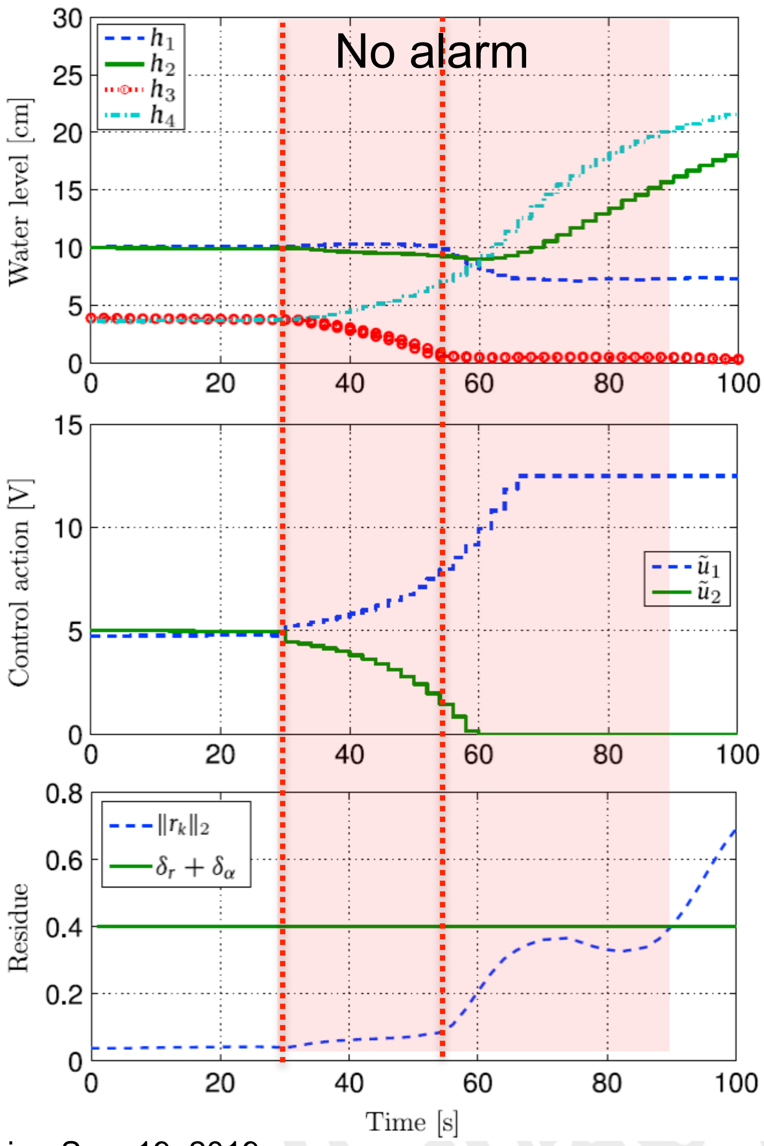
Can the controller/detector always detect the attack?



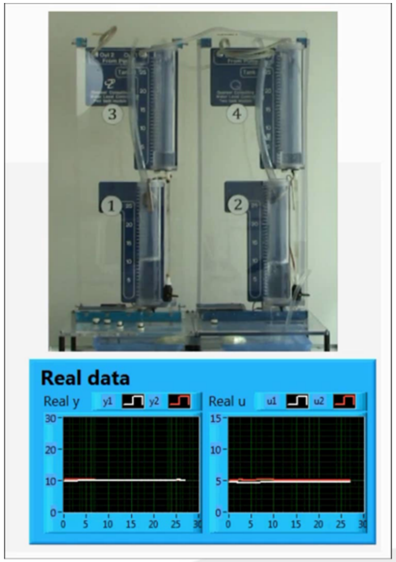
[Teixeira *et al.*, Automatica, 2015]

Undetectable Water Tank Attack [Movie]





Water Tank Model Analysis



- Transfer function matrix from attack to sensor signals

$$G_a(z) = C(zI - A)^{-1}B = \begin{pmatrix} \frac{0.0289}{z-0.8076} & \frac{(1.277z+1.182)\cdot 10^{-3}}{z^2-1.784z+0.7928} \\ \frac{(1.356z+1.24)\cdot 10^{-3}}{z^2-1.754z+0.7643} & \frac{0.02954}{z-0.8347} \end{pmatrix}$$

- Poles = $\{0.8076,\ 0.8347,\ 0.9464,\ 0.9498\}$
- Invariant zeros = $\left\{0.8675,\ 1.0362\right\}$ $\Rightarrow$ Non-minimum phase system
- Applied attack signal (small ϵ)

$$a(k) = 1.0362^k \begin{pmatrix} 0.2281\epsilon \\ -0.2281\epsilon \end{pmatrix}, \quad x_0 = \begin{pmatrix} 0 & 0 & -0.6521\epsilon & 0.6876\epsilon \end{pmatrix}^T$$

satisfies **zero dynamics** and is **masked by** system transient:

$$0 = y(k) = CA^k x_0 + (g_a * a)(k), \quad k \geq 0$$

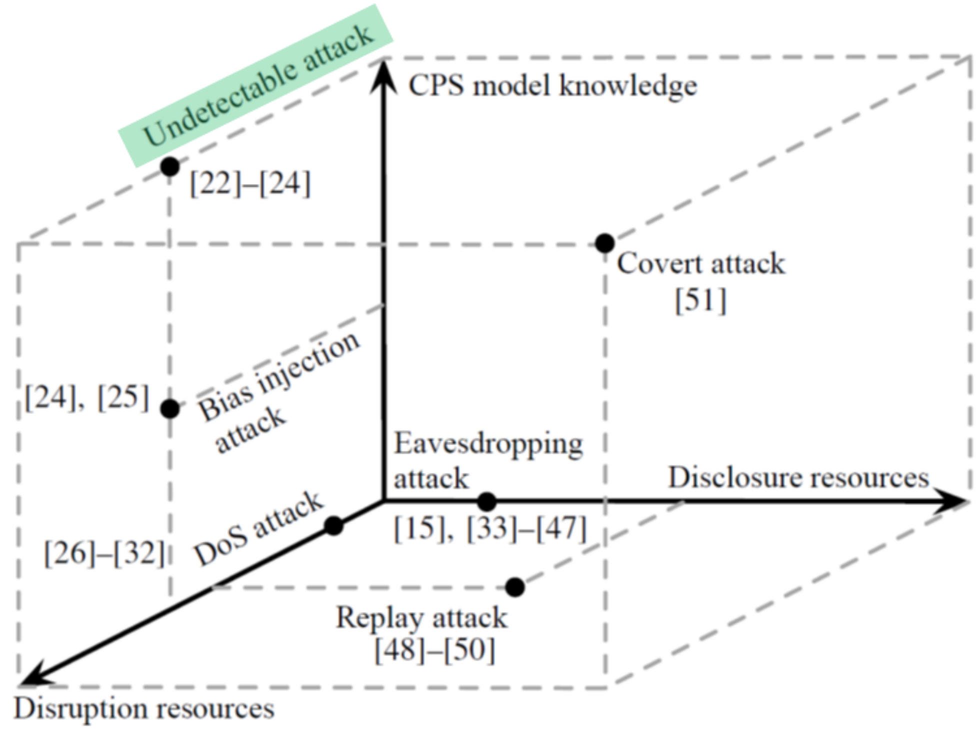
Undetectable Water Tank Attack

2 hacked actuators (u_1 and u_2 = disruption resources)

2 healthy sensors (y_1 and $y_2 \neq$ disruption or disclosure resources)

Can the controller/detector always detect the attack?

Not against an adversary with full CPS model knowledge



Undetectable Attacks and Invariant Zeros

- The Rosenbrock system matrix: $P(z) = \begin{bmatrix} A - zI & B_d & B_a \\ C & D_d & D_a \end{bmatrix}$

Theorem 1: Attack signal $a(k) = z_0^k a_0$, $0 \neq a_0 \in \mathbb{C}^m$, $z_0 \in \mathbb{C}$, is *undetectable* iff there exists $x_0 \in \mathbb{C}^n$ and $d_0 \in \mathbb{C}^o$ such that

$$P(z_0) \begin{bmatrix} x_0 \\ d_0 \\ a_0 \end{bmatrix} = 0$$

- Routine invariant zero computation **(**MATLAB: tzero**)**

[Pasqualetti et al, TAC, 2013], [Sandberg and Teixeira, SoSCYPS, 2016]

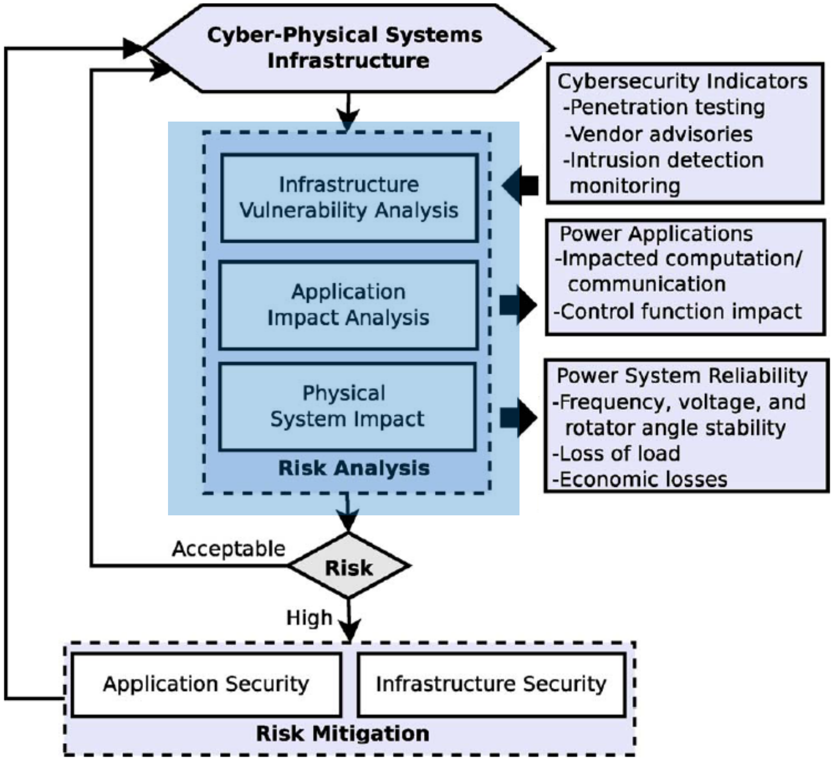
Risk Management Cycle

Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Objectives

- **Risk Analysis**
  - **Likelihood Assessment**
  - **Impact Assessment**

- Risk Mitigation
  - Prevention, Detection, Treatment



[Sridhar *et al.*, Proc. IEEE, 2012]

Tools for Likelihood Assessment: Security Index

$$\alpha_i := \min_{|z_0|\geq 1, x_0, d_0, a_0^i} \|a_0^i\|_0 \quad \text{subject to} \quad P(z_0)\begin{bmatrix} x_0 \\ d_0 \\ a_0^i \end{bmatrix} = 0$$

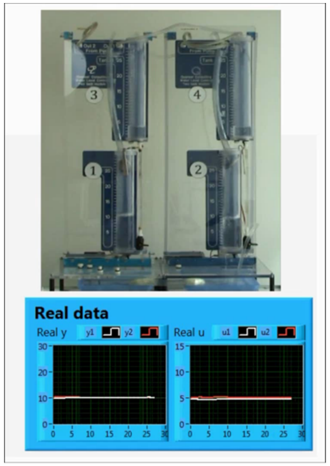
Notation: $\|a\|_0 := |\,\mathrm{supp}(a)\,|$, a^i vector a with i-th element non-zero

Interpretation:

- [Attacker *persistently* targets signal component a_i (condition $|z_0| \geq 1$)]
- α_i is smallest number of attack signals that need to be simultaneously accessed to stage an **undetectable attack** against component a_i
- Estimate likelihood for attack against component i by $\sim 1/\alpha_i$
- Problem NP-hard, but easy when geometric multiplicities of zeros are 1

[Sandberg and Teixeira, SoSCYPS, 2016]

Security Index: Water Tank Example



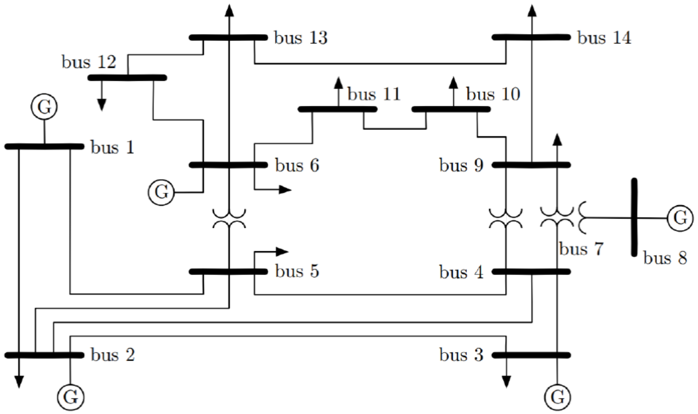
$$G_a(z) = C(zI - A)^{-1}B = \begin{pmatrix} \frac{0.0289}{z-0.8076} & \frac{(1.277z+1.182)\cdot 10^{-3}}{z^2-1.784z+0.7928} \\ \frac{(1.356z+1.24)\cdot 10^{-3}}{z^2-1.754z+0.7643} & \frac{0.02954}{z-0.8347} \end{pmatrix}$$

- Invariant zeros = $\{0.8675,\ 1.0362\} \Rightarrow$ Non-minimum phase system

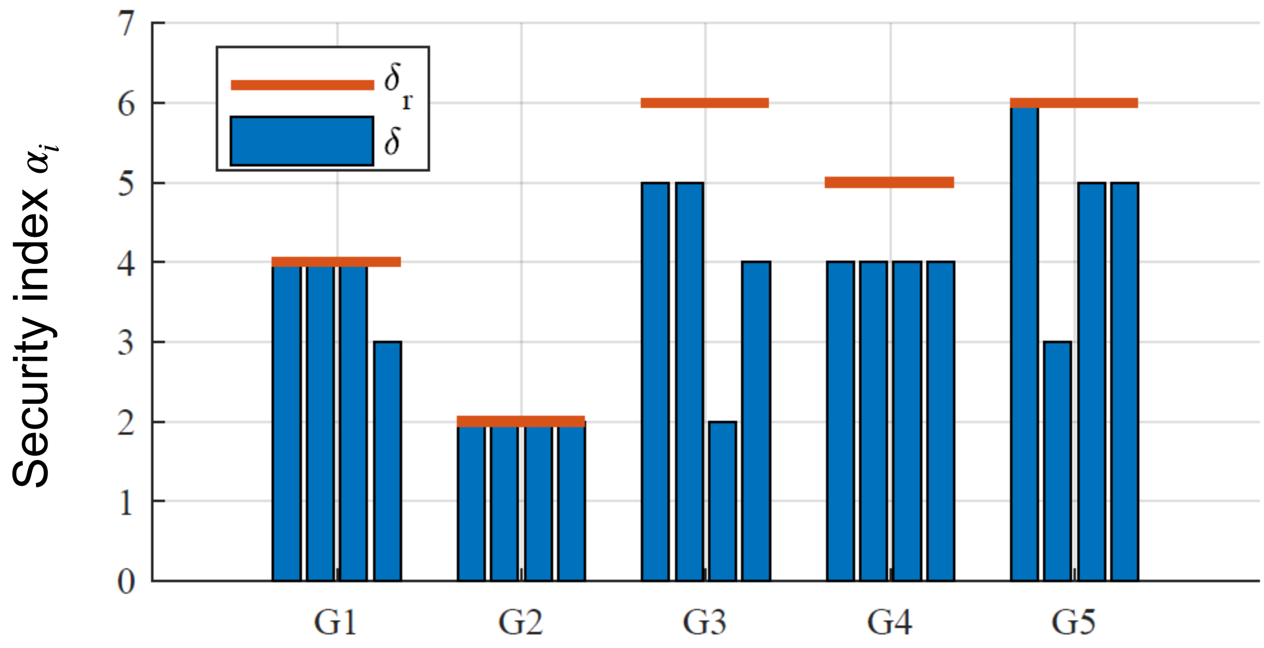
- Persistent undetectable attack: $a(k) = 1.0362^k \begin{pmatrix} 0.2281\epsilon \\ -0.2281\epsilon \end{pmatrix}$

- Only one signal satisfies α_i constraint!
  $\left\|a(k)\right\|_0 = 2 \Rightarrow \alpha_{1,2} = 2$

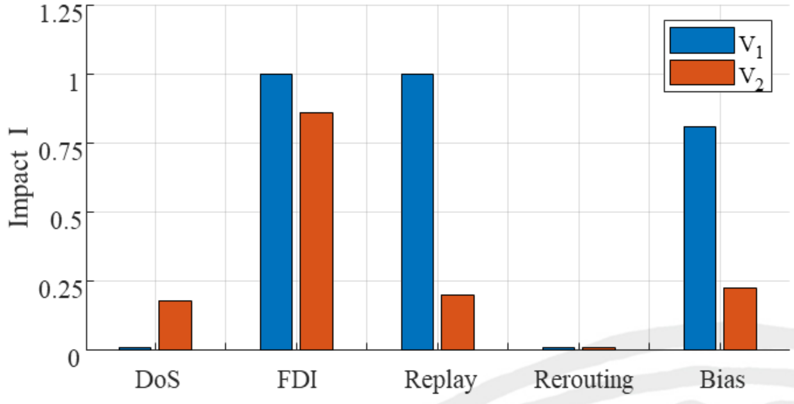
Security Index
IEEE 14 Bus System



[Milosevic *et al.*, arXiv preprint, 2019]



Tools for Impact Assessment: Constrained Reachability



System model:

$$x(k+1) = Ax(k) + B_a a(k)$$
$$y(k) = Cx(k) + D_a a(k)$$

Consider trajectories from k=0, … N.

Impact assessment problem:

$$\begin{array}{ll} \text{maximize}_a & \|x\|_\infty \\ \text{subject to} & a \in \{\text{DoS, Data Injection, Re-routing, Replay, Bias}\} \text{ attack} \\ & y \text{ generates no alarm in } \{\chi^2,\ \text{CUSUM, MEWMA}\} \text{ detector} \end{array}$$

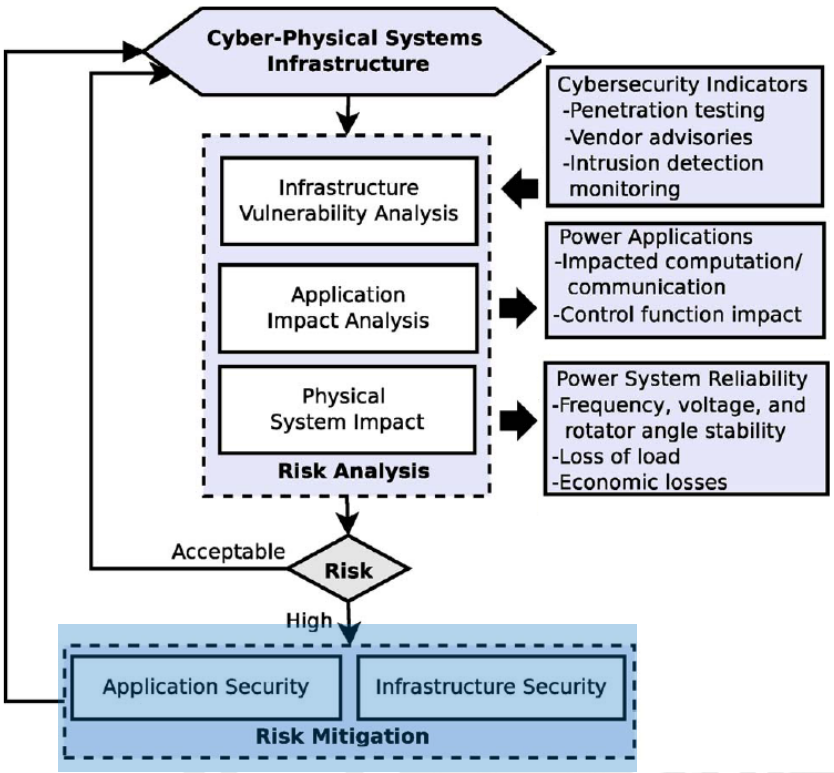
Theorem 3: [Milosevic *et al.*, ECC, 2018]

i. Constraints are convex
ii. Optimal value found by solving set of convex optimization problems

Risk Management Cycle

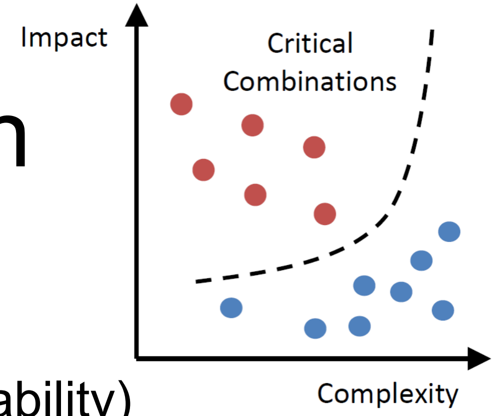
Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Obje
- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment
- **Risk Mitigation**
  - Prevention, Detection, Treatment



[Sridhar *et al.*, Proc. IEEE, 2012]

Tools for Risk Mitigation



- **Prevention** (decrease likelihood by reducing vulnerability)
  - Watermarking and Moving Target Defense
  - Coding and Encryption Strategies
  - Rational Security Allocation
  - Privacy-preservation by Noise Injection
- **Detection** (continuous anomaly monitoring)
  - Tuning of Detector Thresholds
  - Secure State Estimation
  - Watermarking and Moving Target Defense
  - Distributed Algorithms
  - Methods Related to Robust Statistics
- **Treatment** (compensate for or neutralize detected attack)
  - Secure State Estimation
  - Countering DoS Attacks
  - Distributed Algorithms
  - Methods Related to Robust Statistics

[Chong *et al.*, ECC, 2019]

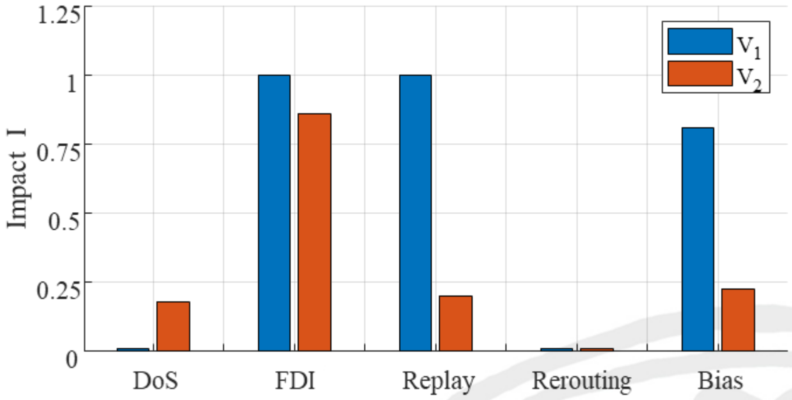
Outline

- Motivation

- Risk management
  - Scenario characterisation
  - Risk Analysis
  - Risk Mitigation

- **Secure control: from analysis to design**
  - Problem formulation
  - Metrics in Fault-Tolerant Control
  - Metrics in Secure Control
    - Computation and design problem

Tools for Impact Assessment: Constrained Reachability



System model:

$$x(k+1) = Ax(k) + B_a a(k)$$
$$y(k) = Cx(k) + D_a a(k)$$

Consider trajectories from k=0, … N.

Impact assessment problem:

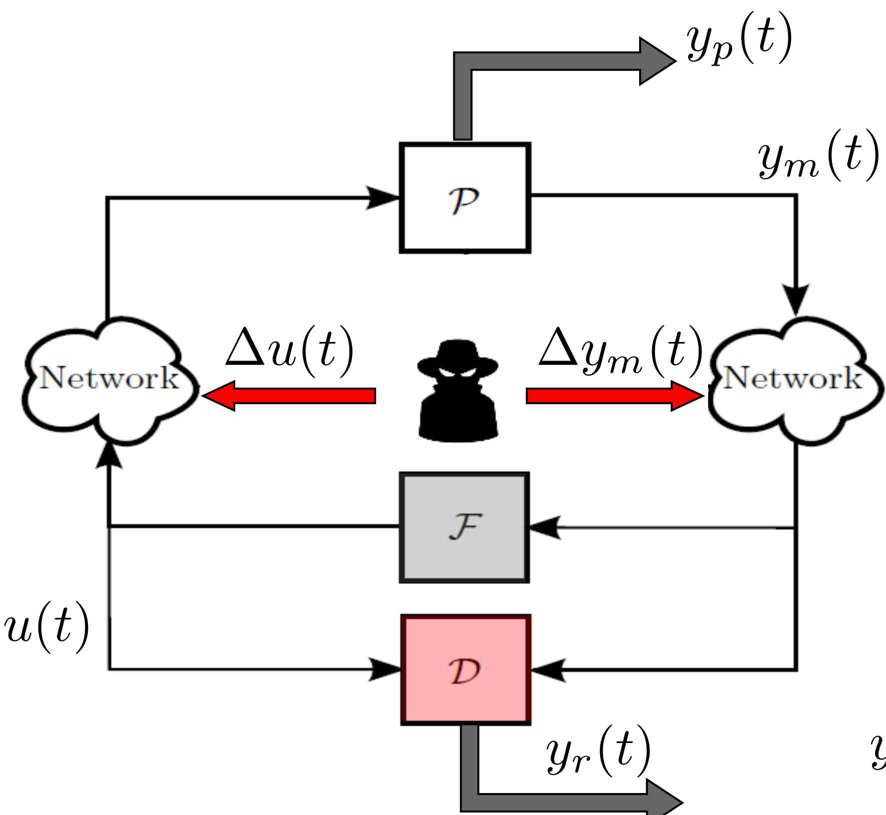
$$\begin{aligned} \text{maximize}_a \quad & \|x\|_\infty \\ \text{subject to} \quad & a \in \{\text{DoS, Data Injection, Re-routing, Replay, Bias}\} \text{ attack} \\ & y \text{ generates no alarm in } \{\chi^2,\ \text{CUSUM, MEWMA}\} \text{ detector} \end{aligned}$$

Drawbacks:

Only looks at the terminal state

It only serves for analysis, not for design…

Networked Control System under Attack

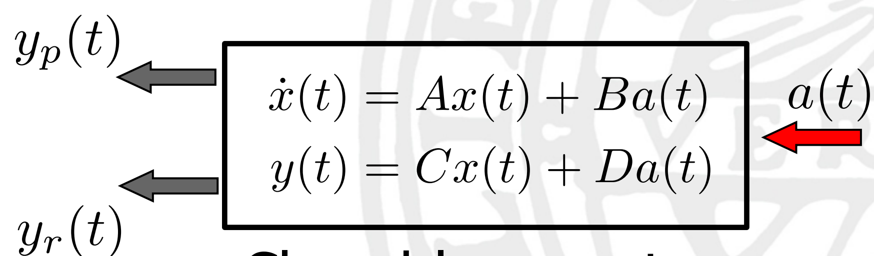


$\|y_r\|^2_{\mathcal{L}_2} \geq 1 \Rightarrow$ **Alarm!**

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

- Performance output: $y_p(t)$ (e.g. physical state)
- Detector output: $y_r(t)$

- Actuator and Sensor data corruption: $a(t) = \begin{bmatrix} \Delta u(t) \\ \Delta y_m(t) \end{bmatrix}$



Closed-loop system

Adversary Model

$y_p(t)$

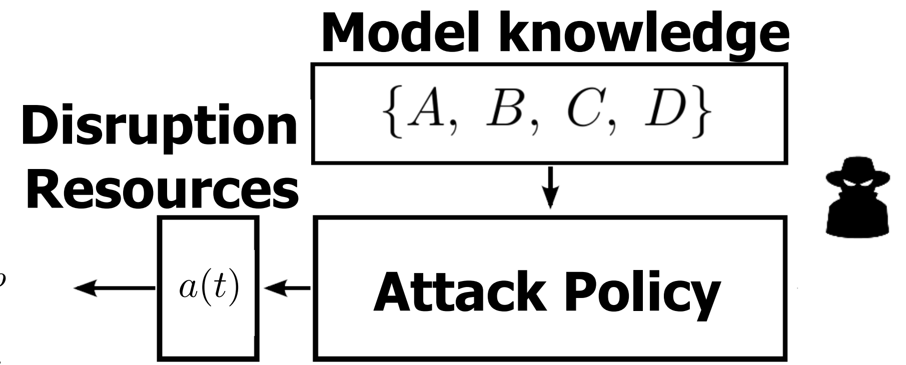
$$\dot{x}(t) = Ax(t) + Ba(t)$$
$$y(t) = Cx(t) + Da(t)$$

$a(t)$

$y_r(t)$

Closed-loop system

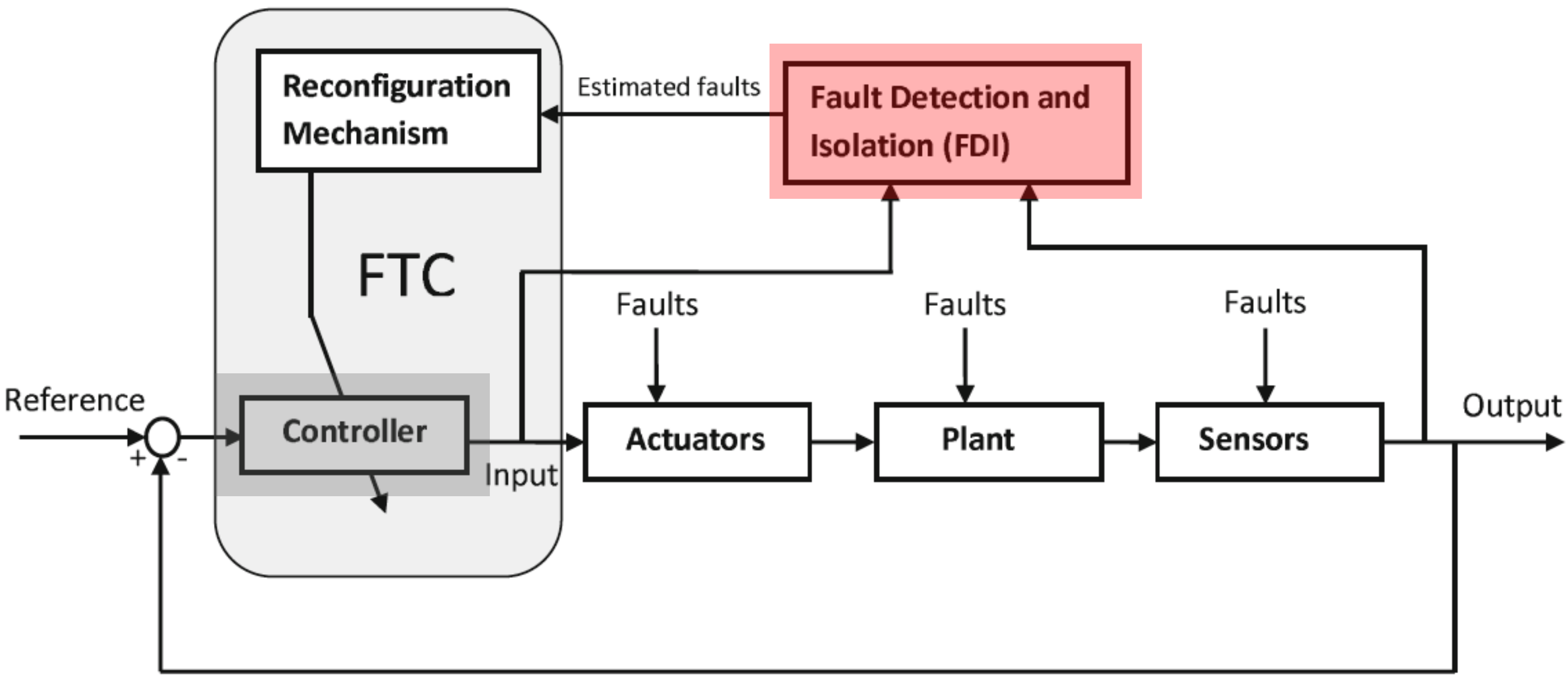
$$\Sigma_p = (A, B, C_p, D_p),\ G_p(s) = C_p(sI - A)^{-1}B + D_p$$
$$\Sigma_r = (A, B, C_r, D_r),\ G_r(s) = C_r(sI - A)^{-1}B + D_r$$



- **Model knowledge:** Dynamical model of the closed-loop system
- **Disruption resources:** (Small no. of) measurement and actuation channels
- **Attack policy:** Maximise the impact on performance without raising alarms

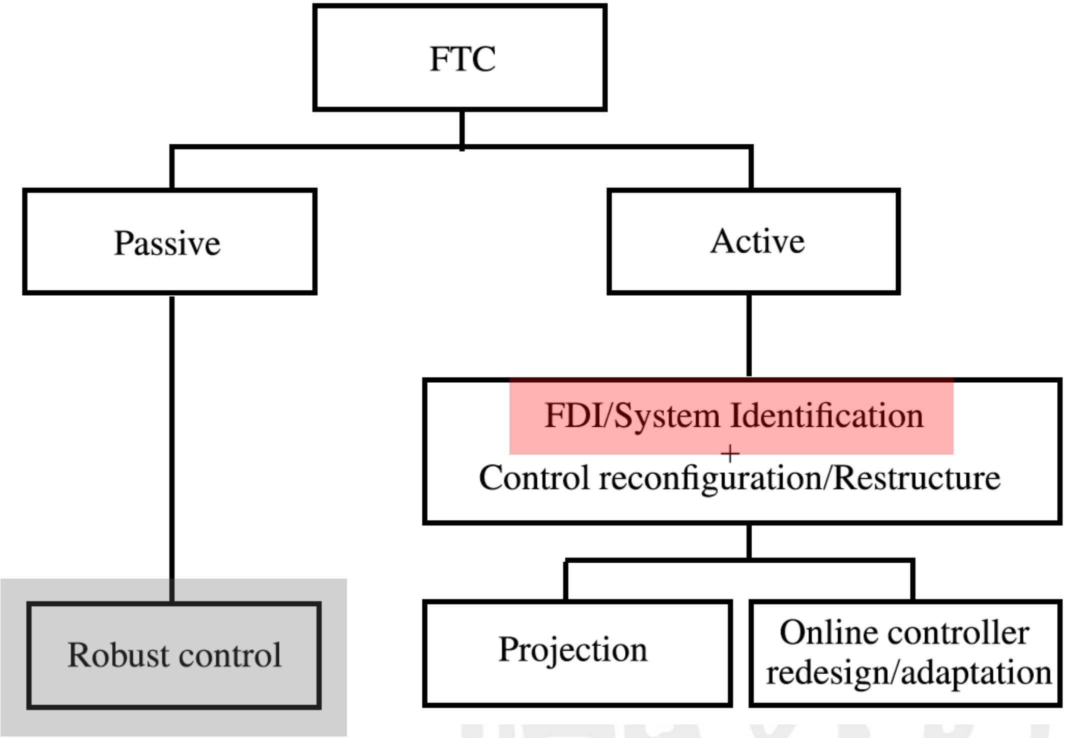
How to analyse and design the system with such an attack?

Fault-Tolerant Control



Fault-Tolerant Control

- While fault is not detected:
  - Robust controller
- Once fault is detected:
  - Switch/adapt controller



FTC design objectives

- **Robust control:** find a controller that
  - Minimizes the "worst-case" (largest) **impact** of unit-energy faults
  - i.e.: optimal "H infinity" control
- **Fault detection:** find an observer/filter that
  - Maximizes the "worst-case" (smallest) **detectability** of unit-energy faults
  - i.e.: optimal "H minus" (H_) filter design
- Both are based on *sensitivity metrics*:
  - H-inf norm: largest **impact** of <u>unit-energy</u> faults
  - H_ index: smallest **detectability** of <u>unit-energy</u> faults

Classical Sensitivity Metrics (1)

$y_p(t)$

$$\dot{x}(t) = Ax(t) + Ba(t)$$
$$y(t) = Cx(t) + Da(t)$$

$a(t)$

$y_r(t)$

Closed-loop system

$\mathcal{L}_{2e}$ = "signals with finite energy over finite time intervals"

$$\|y\|_{\mathcal{L}_2}^2 \triangleq \int_{-\infty}^{+\infty} \|y(t)\|_2^2 \, dt$$

- H-inf norm:

$$\begin{aligned} \gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_{2e}} \quad & \|y_p\|_{\mathcal{L}_2} \\ \text{s.t.} \quad & \|a\|_{\mathcal{L}_2} = 1 \\ & x(0) = 0 \end{aligned}$$

- Frequency Domain:

$$\gamma_{H_\infty} = \sup_{w \geq 0} \quad \bar{\sigma}_p(jw)$$

$$\begin{aligned} \bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^{n_a}} \quad & \|G_p(s)a\|_2 \\ \text{s.t.} \quad & \|a\|_2 = 1 \end{aligned}$$

- H_ index:

$$\begin{aligned} \gamma_{H_-} \triangleq \inf_{a \in \mathcal{L}_{2e}} \quad & \|y_r\|_{\mathcal{L}_2} \\ \text{s.t.} \quad & \|a\|_{\mathcal{L}_2} = 1 \\ & x(0) = 0 \end{aligned}$$

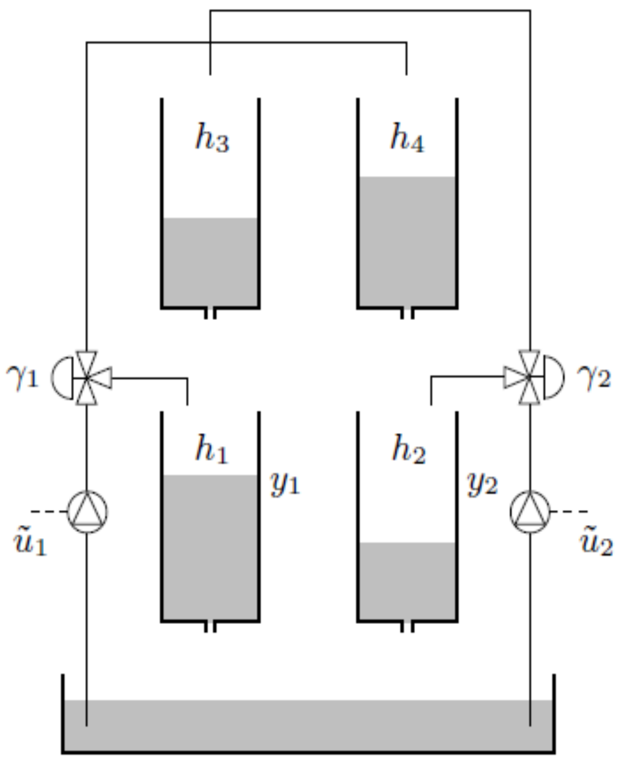
- Frequency Domain:

$$\gamma_{H_-} = \inf_{w \geq 0} \quad \underline{\sigma}_r(jw)$$

$$\begin{aligned} \underline{\sigma}_r(s) = \inf_{a \in \mathbb{C}^{n_a}} \quad & \|G_r(s)a\|_2 \\ \text{s.t.} \quad & \|a\|_2 = 1 \end{aligned}$$

Example: closed-loop system

- **Process:** Quadruple tank
  - (non-minimum phase setup, one unstable zeros from u to y)
- **Controller:** LQG with integral action
  - Performance output = plant's states
    $$y_p(t) = x_p(t)$$
- **Detector:** LQG's Kalman filter
  - Residual = output estimation error
    $$y_r(t) = y(t) - \hat{y}(t)$$
- The adversary is able to corrupt Actuator 1 (u_1)



Classical Sensitivity Metrics (2)

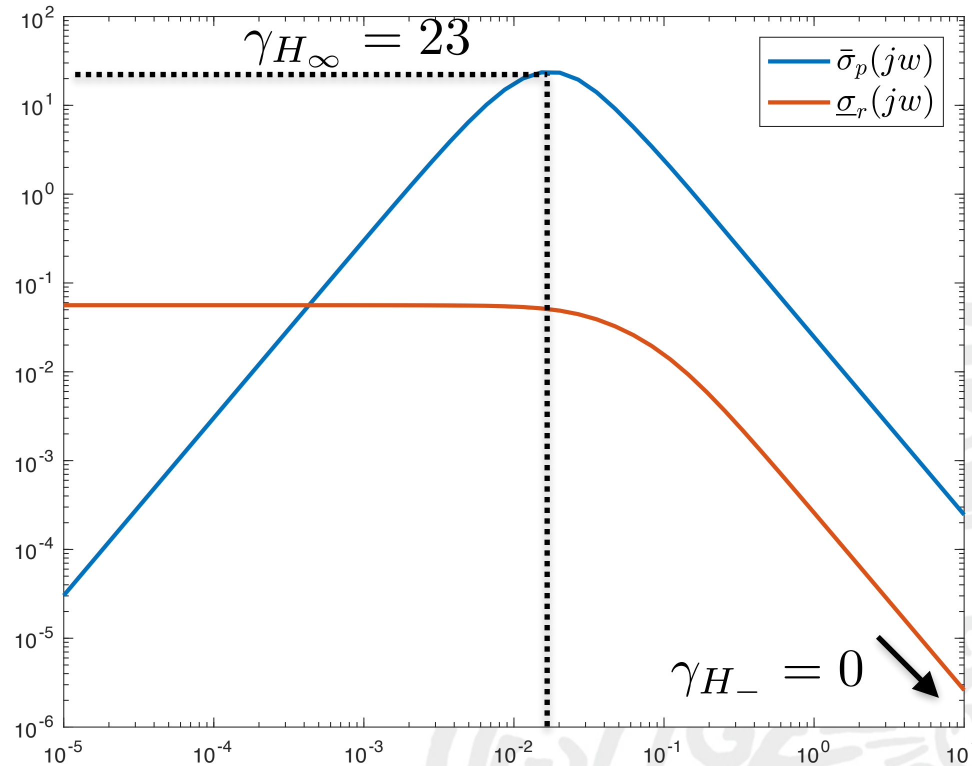
- H-inf norm

$$\gamma_{H_\infty} = \sup_{w \geq 0} \quad \bar{\sigma}_p(jw)$$

- H_ index

$$\gamma_{H_-} = \inf_{w \geq 0} \quad \underline{\sigma}_r(jw)$$

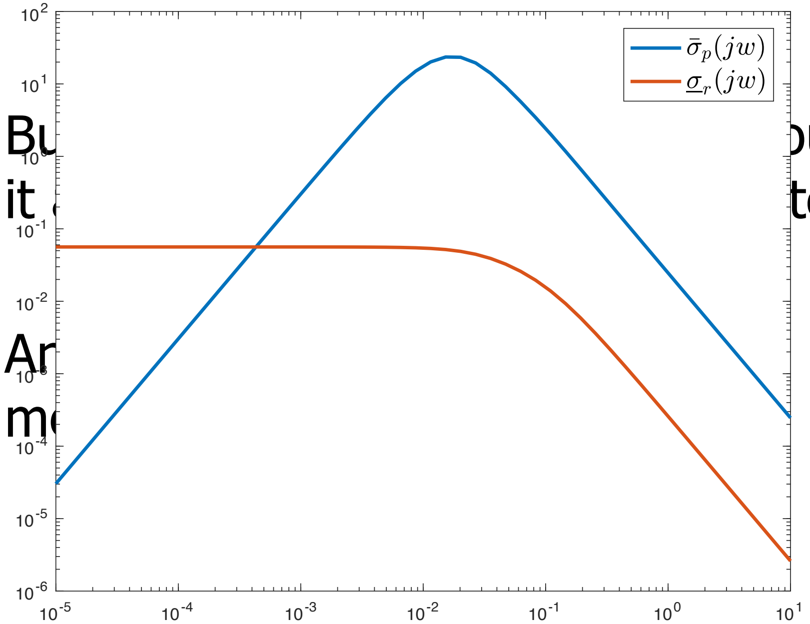
- Least detectable fault has little impact…



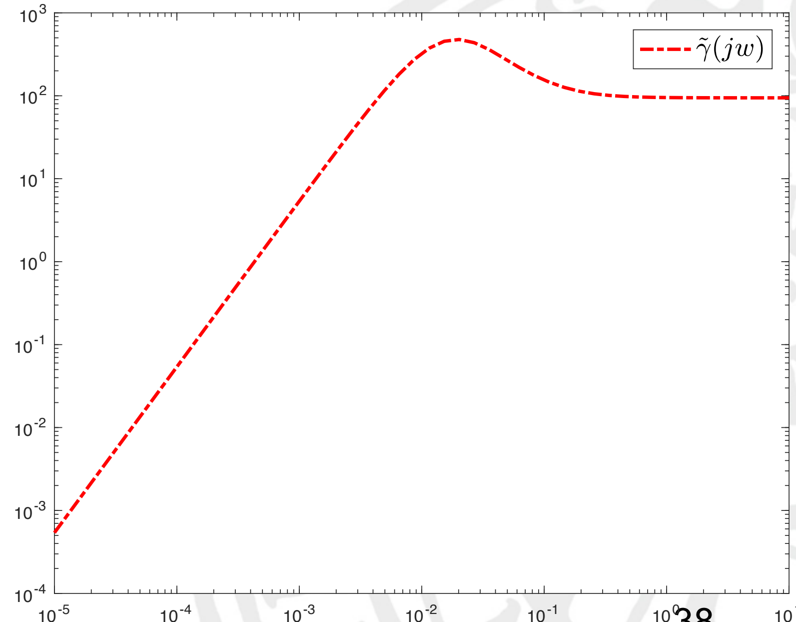
- **Limitation:** worst-case frequency is not the same
  - Means each metric looks at **different** worst-case inputs!

Sensitivity metric for security of control systems (1)

- **Attack policy:** <u>Maximise</u> the impact on <u>performance</u> <u>without raising alarms</u>
  - Requires a <u>combination</u> of **impact on performance** and **detection!**
- "Desired" metric: <u>combine</u> **H-inf** and **H_:** $\tilde{\gamma}^\star = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$



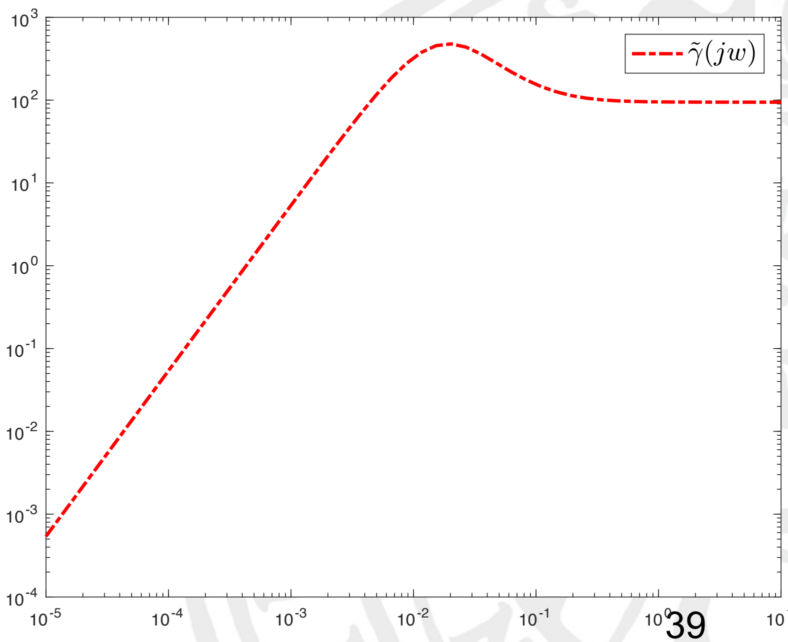
- Bu ould it te?
- Ar m



Sensitivity metric for security of control systems (1)

- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**
- "Desired" metric: combine **H-inf** and **H_:** $\tilde{\gamma}^\star = \sup_{w\geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

- But this is an heuristic… When would it actually be meaningful / accurate?
- And how do we formalise such a metric?
  - … coming next!



Sensitivity metric for security of control systems (2)

- **Attack policy:** Maximise the impact on performance without raising alarms
  - Maximize y_p, while keeping y_r small — **Output-to-ouput gain**:

$$\begin{aligned} \gamma^\star \triangleq \sup_{a \in \mathcal{L}_{2e}} \quad & \|y_p\|_{\mathcal{L}_2} \\ \text{s.t.} \quad & \|y_r\|_{\mathcal{L}_2} = 1 \\ & x(0) = 0 \end{aligned}$$

[Teixeira et al., CDC 15]

- Note: Input is not directly constrained (may be exponentially increasing)

$\mathcal{L}_{2e} =$ "signals with finite energy over finite time intervals"

- Imposes a gain constraint on the outputs:

$$\gamma^{\star^2} \|y_r\|^2_{\mathcal{L}_2} \geq \|y_p\|^2_{\mathcal{L}_2}$$

- An equivalent formulation:

$$\begin{aligned} \gamma^{\star^2} = \min_{\beta \geq 0} \quad & \beta \\ \text{s.t.} \quad & \beta \|y_r\|^2_{\mathcal{L}_2} - \|y_p\|^2_{\mathcal{L}_2} \geq 0, \ \forall\, a \in \mathcal{L}_{2e}, \ x(0) = 0 \end{aligned}$$

Computation: Linear Matrix Inequality

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$

$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- The value of the metric can be obtained by solving this convex problem
- Controller / Detector design is possible
  - A, B, C, D depend linearly on the controller / detector
  - … but leads to Bilinear Matrix Inequalities
  - Exploit structure to convexify the problem - ongoing work
- Other metrics can be formulated in a similar fashion [Teixeira, CDC 19]

UPPSALA UNIVERSITET

Summary

- Secure Control systems - a risk management approach

- Classical sensitivity metrics are not adequate for security-related problems / malicious adversary models
  - They focus **either** on impact or on detection **separately**

- Output-to-output gain captures **both** impact and detection
  - Maximize energy of ”cost signal”; While keeping ”anomaly detector signal” small
  - Computation based on LMIs
  - Controller / Detector design through BMIs

andre.teixeira@angstrom.uu.se

Thank you!

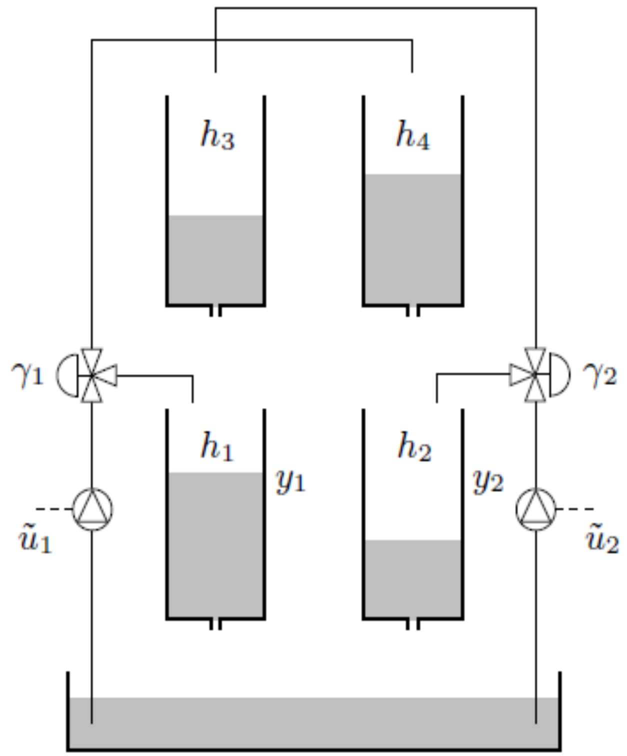
References

- [Kaplan & Garrick, 1981] S. Kaplan and B. J. Garrick, "On The Quantitative Definition of Risk", Risk Analysis, vol. 1, issue 1, 11-27, 1981.
- [Bishop, 2002] M. Bishop, "Computer Security: Art and Science", Addison-Wesley Professional, 2002
- [Sridhar *et al.*, Proc. IEEE, 2012] S. Sridhar, A. Hahn, and M. Govindarasu. "Cyber-physical system security for the electric power grid". *Proceedings of the IEEE*, 100(1), 210-224, 2012.
- [Teixeira *et al.*, Automatica, 2015] **A.M.H. Teixeira**, I. Shames, H. Sandberg, and K. H. Johansson. "A Secure Control Framework for Resource-Limited Adversaries". Automatica, vol. 51, pp. 135-148, 2015.
- [Teixeira *et al.*, CSM 2015] **A.M.H. Teixeira**, K. C. Sou, H. Sandberg, and K. H. Johansson. "Secure Control Systems: A Quantitative Risk Management Approach". IEEE Control System Magazine, vol. 35, no. 1, pp. 24-25, Feb. 2015.
- [Chong *et al.*, ECC, 2019] M. Chong, H. Sandberg, **A.M.H. Teixeira**. "A Tutorial Introduction to Security and Privacy for Cyber-Physical Systems", European Control Conference, Napole, Italy, 2019.
- [Pasqualetti *et al.*, TAC, 2013] F. Pasqualetti, F. Dörfler, F. Bullo, "Attack Detection and Identification in Cyber-Physical Systems". IEEE Transactions on Automatic Control 58(11):2715-2729, 2013.
- [Sandberg and Teixeira, SoSCYPS, 2016] H. Sandberg and **A.M.H. Teixeira**. "From control system security indices to attack identifiability". In Proc. 2016 Science of Security for Cyber-Physical Systems Workshop (SOSCYPS), 2016.
- [Milosevic *et al.*, arXiv preprint, 2019] J. Milosevic, **A.M.H. Teixeira**, H. Sandberg, K. H. Johansson. "Actuator Security Indices Based on Perfect Undetectability: Computation, Robustness, and Sensor Placement". Under Review, available on ArXiv.
- [Milosevic *et al.*, ECC 2018] J. Milosevic, D. Umsonst, H. Sandberg, K. H. Johansson, "Quantifying the Impact of Cyber-Attack Strategies for Control Systems Equipped with an Anomaly Detector", European Control Conference 2018.
- [Teixeira, CDC 19] **A.M.H. Teixeira**, "strictly proper systems", *2019 58th IEEE Conference on Decision and Control (CDC)*, Nice, France, Dec 2019
- [Teixeira *et al.*, CDC 15] **A.M.H. Teixeira,** H. Sandberg and K. H. Johansson, "Strategic stealthy attacks: The output-to-output l2-gain", *2015 54th IEEE Conference on Decision and Control (CDC)*, Osaka, 2015

Extra slides

Example: Attack scenarios

- The adversary is able to corrupt:
  - A. Actuator 1
  - B. Actuator 1 and Actuator 2 (unstable zero —> y=0)
  - C. Actuator 1 and Sensor 1
  - D. Actuator 1 and Sensor 2
- Evaluate which scenarios are most critical (largest sensitivity)

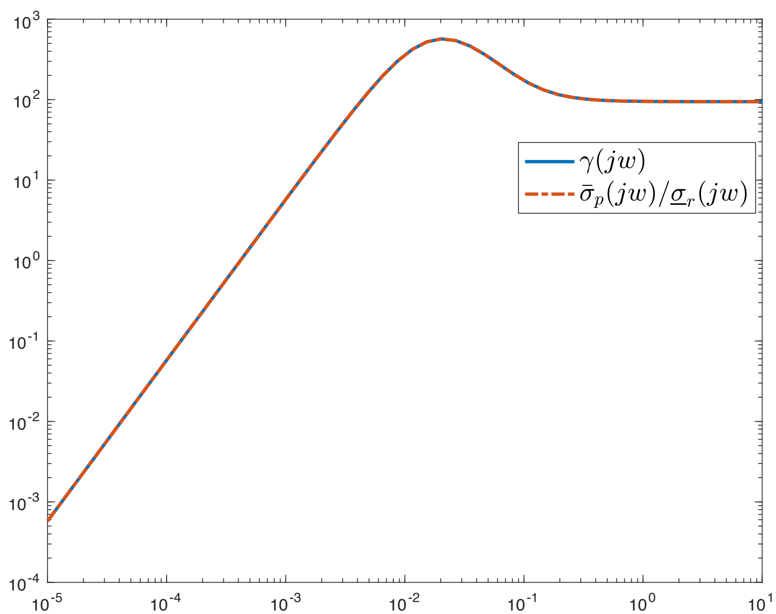


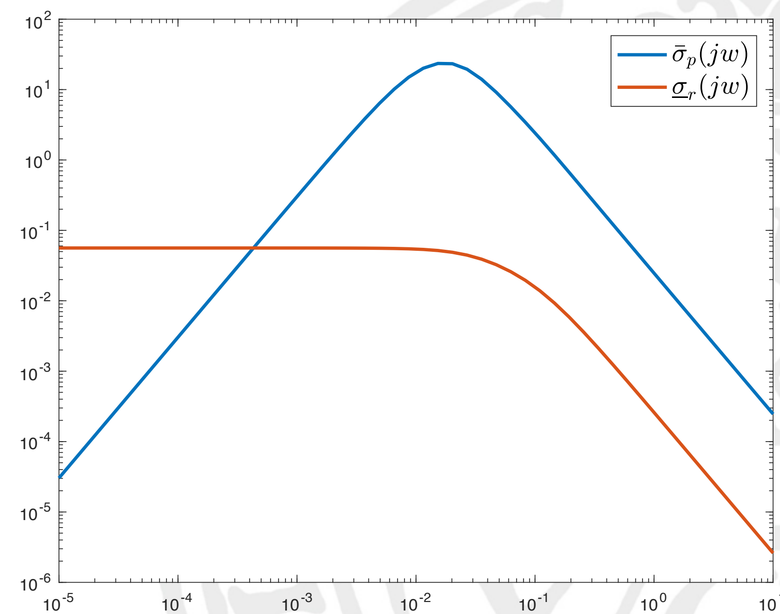
$$\gamma(s) \triangleq \sup_{a \in \mathbb{C}^{n_a}} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2} \qquad \tilde{\gamma}^\star = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

$$\gamma^\star = \sup_{s \in \mathcal{S}} \gamma(s)$$

Example: Actuator 1

- Gain vs heuristic: $\gamma^\star = 23.8, \ \tilde{\gamma}^\star = 23.8$ (LMI gives: $\gamma^\star = 24.1$)
- Gain over Frequency: $\gamma(jw) = \tilde{\gamma}(jw)$ $\quad \sup_{w \geq 0} \gamma(jw) = 23.8$





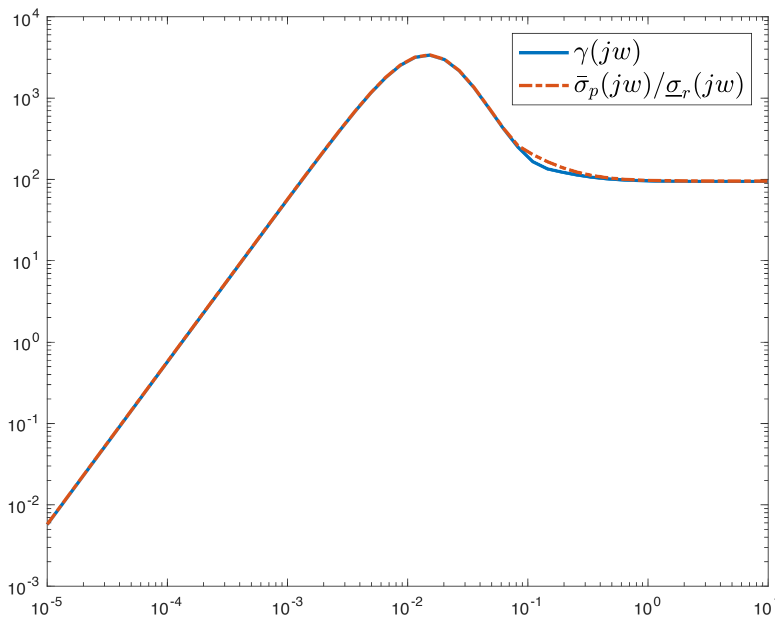
$$\gamma(s) \triangleq \sup_{a \in \mathbb{C}^{n_a}} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2}$$

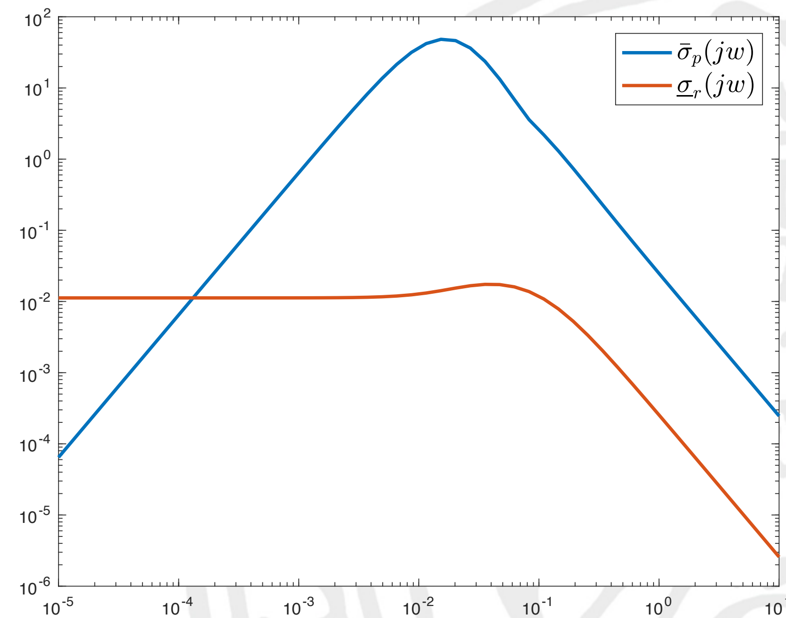
$$\gamma^\star = \sup_{s \in \mathcal{S}} \gamma(s)$$

$$\tilde{\gamma}^\star = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

Example: Actuator 1 and Actuator 2

- Gain vs heuristic: $\gamma^\star = \infty, \ \tilde{\gamma}^\star = 58.3$ (unstable zero)
- Gain over Frequency: $\gamma(jw) \leq \tilde{\gamma}(jw)$ $\quad \sup_{w \geq 0} \gamma(jw) = \tilde{\gamma}^\star$



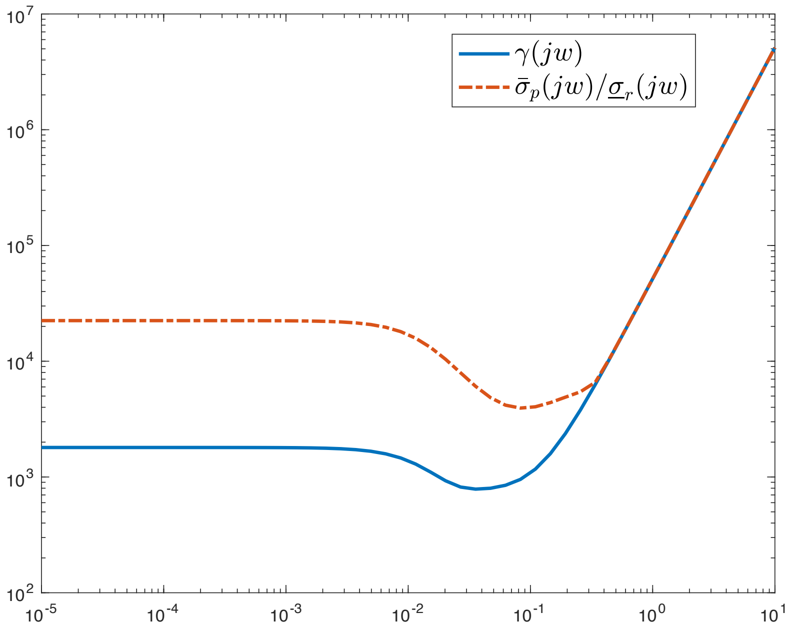


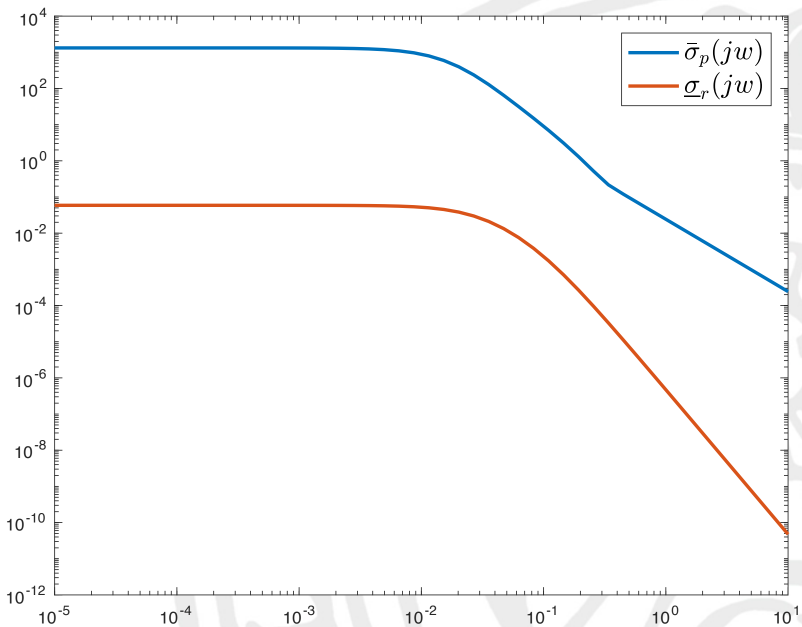
$$\gamma(s) \triangleq \sup_{a\in\mathbb{C}^{n_a}} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2} \qquad \tilde{\gamma}^\star = \sup_{w\geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

$$\gamma^\star = \sup_{s\in\mathcal{S}} \gamma(s)$$

Example: Actuator 1 and Sensor 1

- Gain vs heuristic: $\gamma^\star = \infty, \ \tilde{\gamma}^\star = \infty$ (zeros at infinity)
- Gain over Frequency: $\gamma(jw) \leq \tilde{\gamma}(jw) \qquad \sup_{w\geq 0} \gamma(jw) = \gamma^\star$





$$\gamma(s) \triangleq \sup_{a \in \mathbb{C}^{n_a}} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2}$$

$$\gamma^\star = \sup_{s \in \mathcal{S}} \gamma(s)$$

$$\tilde{\gamma}^\star = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

Example: Actuator 1 and Sensor 2

- Gain vs heuristic: $\gamma^\star = 73.1, \ \tilde{\gamma}^\star = 253.5$ (zeros at infinity are cancelled)
- Gain over Frequency: $\gamma(jw) \leq \tilde{\gamma}(jw)$ $\sup_{w \geq 0} \gamma(jw) = \gamma^\star$