Security Metrics for Control Systems

from Analysis to Design

André Teixeira

Division of Systems and Control

Department of Information Technology

Uppsala University

andre.teixeira@it.uu.se

www.andre-teixeira.eu

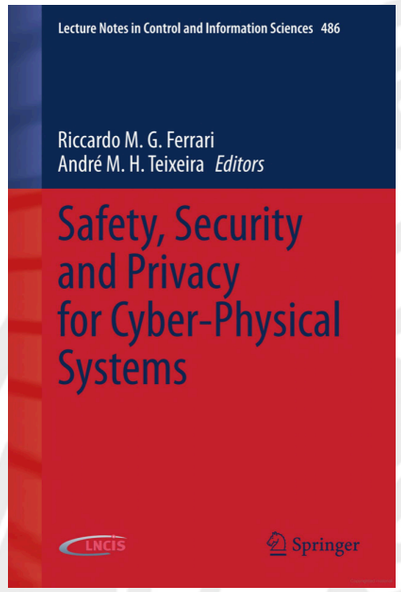
June 27, 2022

Outline

- Cybersecurity in Control Systems
  - Security Game
  - Risk Management
    - Mapping to Control System Design
    - Example: undetectable attacks

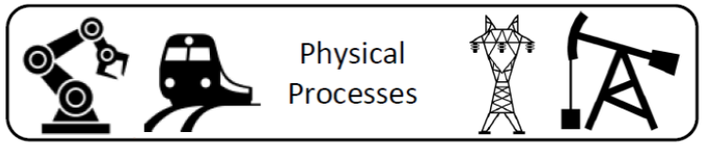
- Security Metrics for Control Systems
  - Classical metrics in control engineering
  - Novel security metric for analysis and design
    - Analysis and design problems
    - Structural limitations - invariant zeros

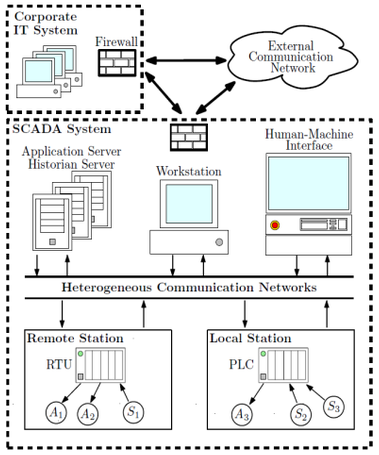
- Additional topics
  - Variations of the adversary models and their respective metrics
  - Incorporating uncertainty for robust open-loop attacks

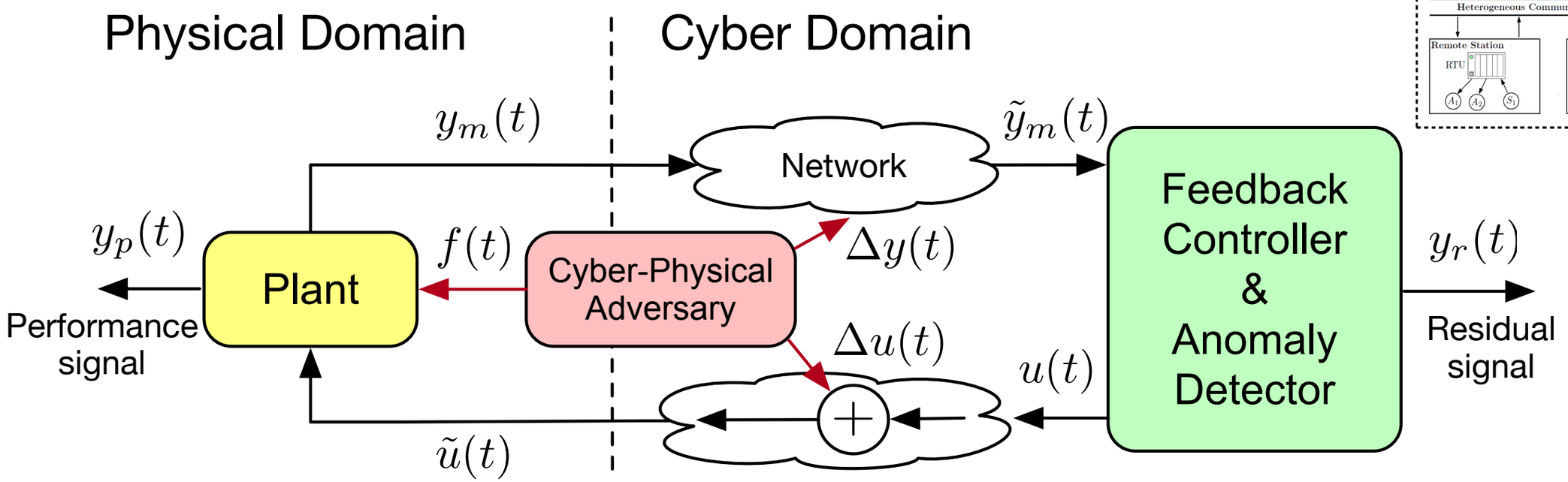


A. M. H. Teixeira, “Security metrics for control systems,” in Safety, Security and Privacy for Cyber-Physical Systems, R. M. Ferrari and A. M. H. Teixeira, Eds. Cham: Springer International Publishing, 2021.

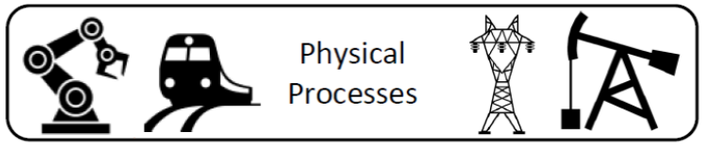
“The Security Game”: key ingredients

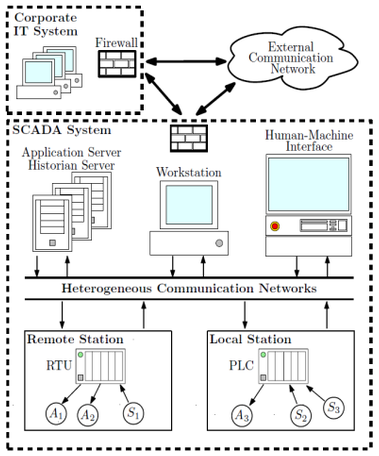


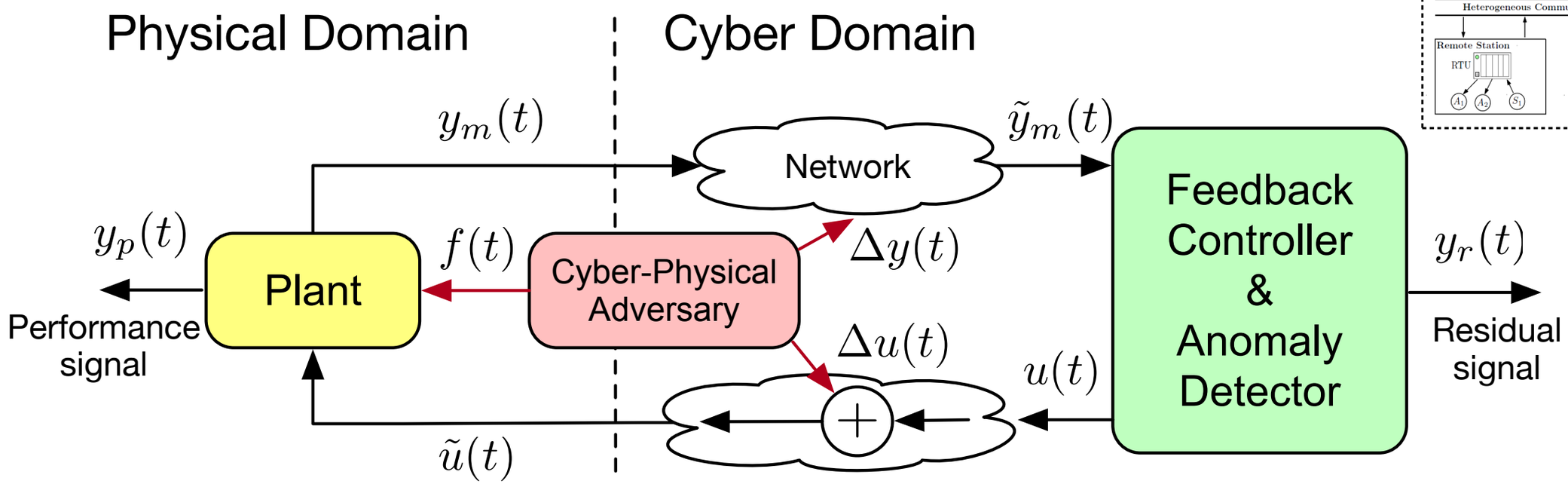




“The Security Game”: key ingredients



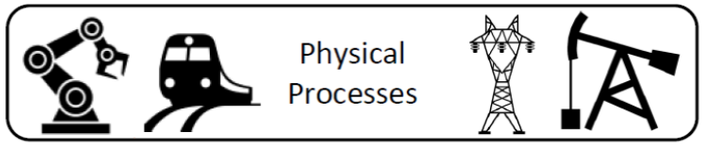


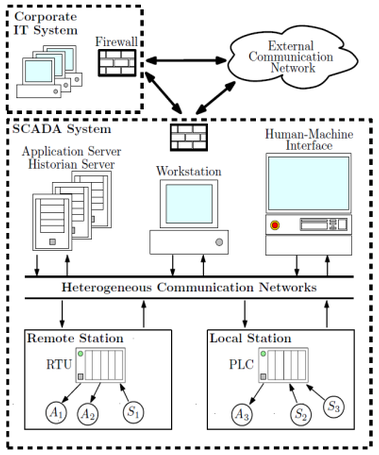


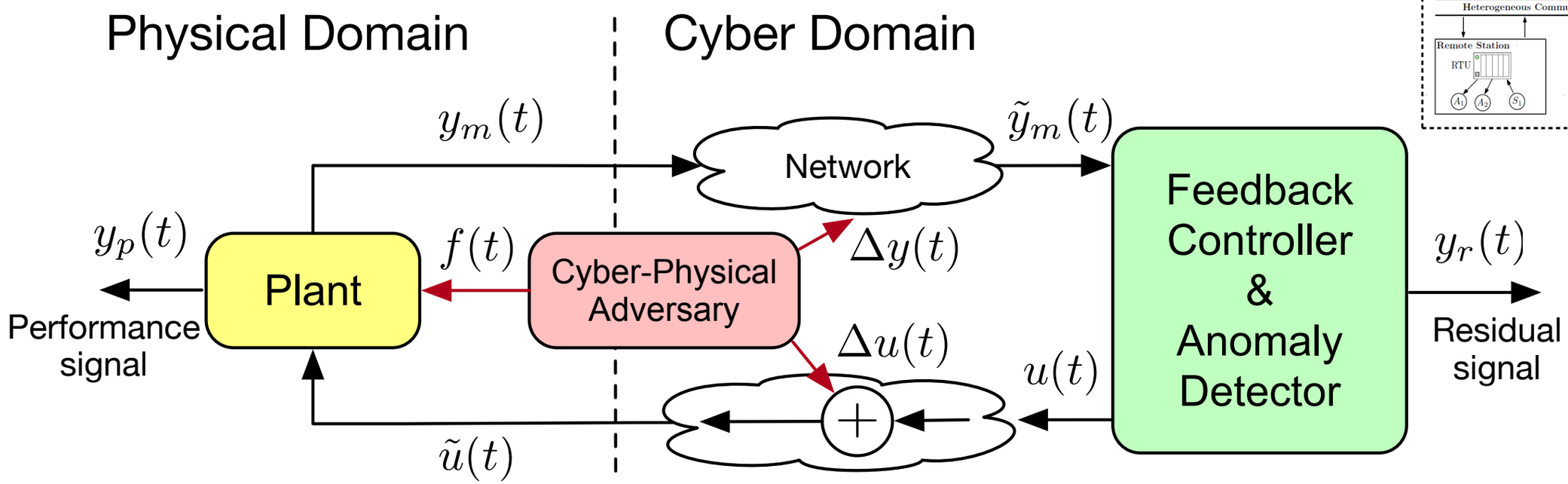
Need tools and strategies to understand and mitigate attacks:

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

“The Security Game”: key ingredients







Need tools and strategies to understand and mitigate attacks:

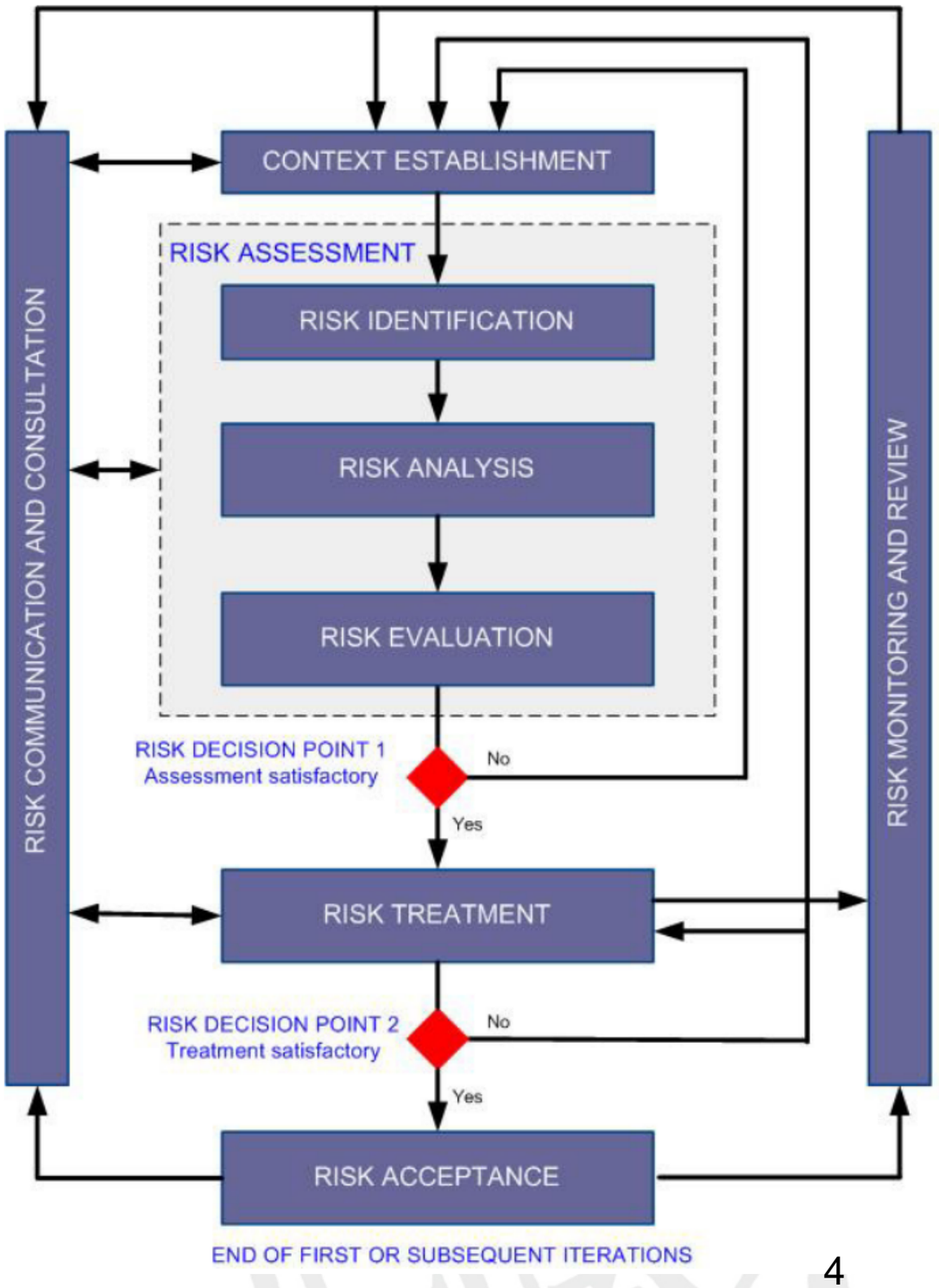
- **Which threats** should we care about?
- **What impact** can we expect from attacks?
- **Which resources** should we **protect**, and how?

- How to find answers: **Risk Management + Control Engineering**

Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]



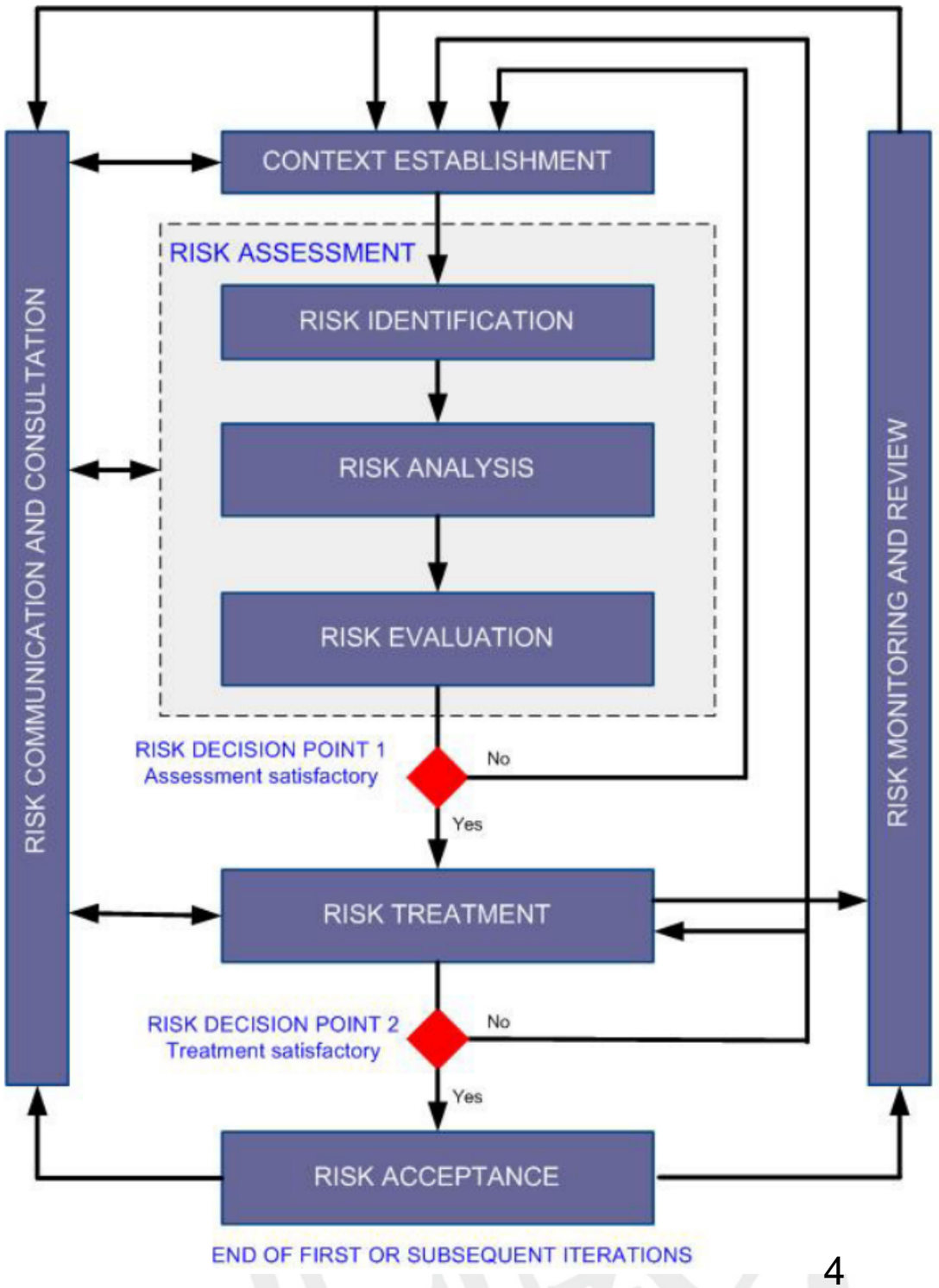
[ISO 31000]

Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management



[ISO 31000]

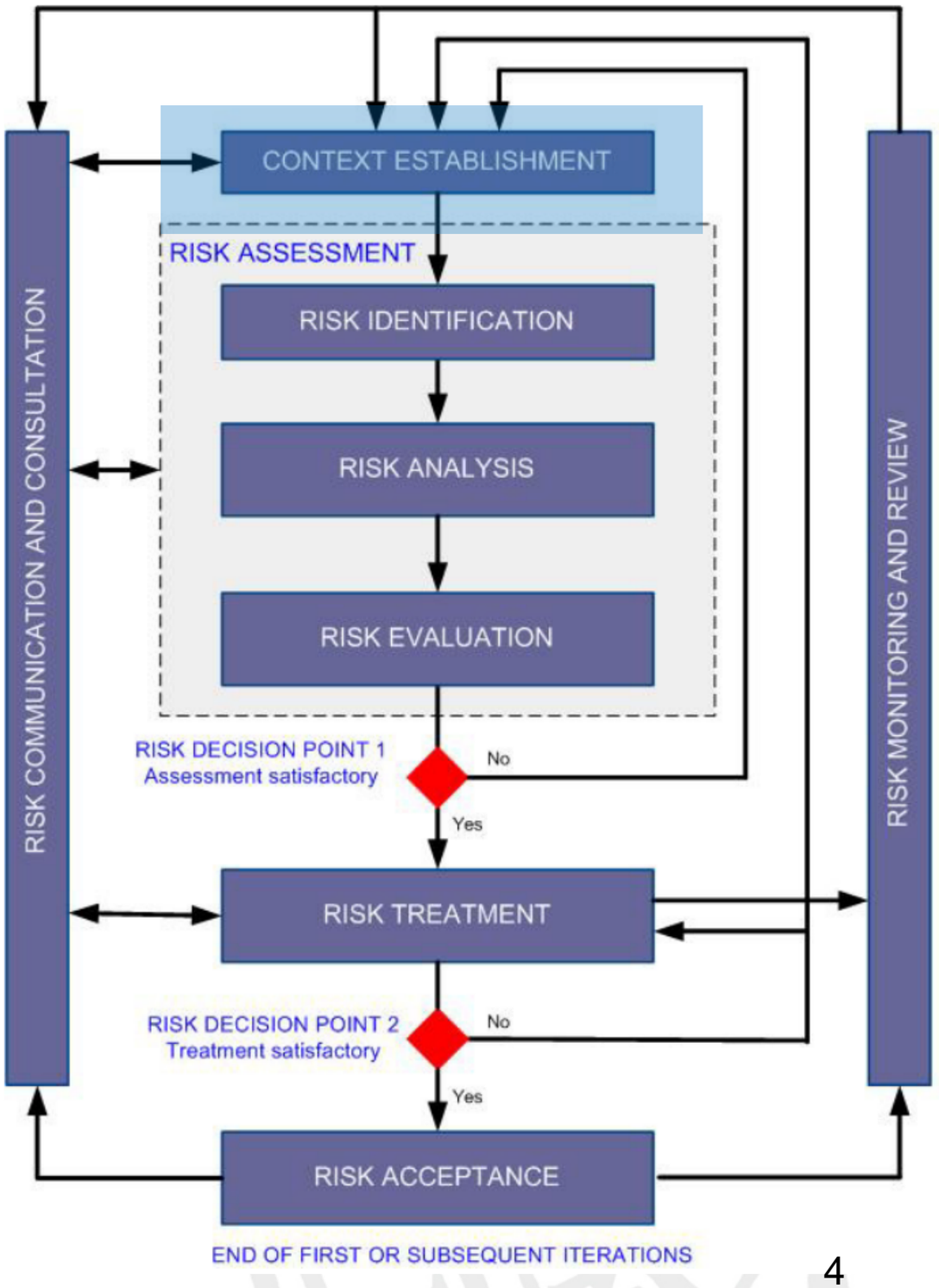
Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Objectives



[ISO 31000]

Risk Management Cycle

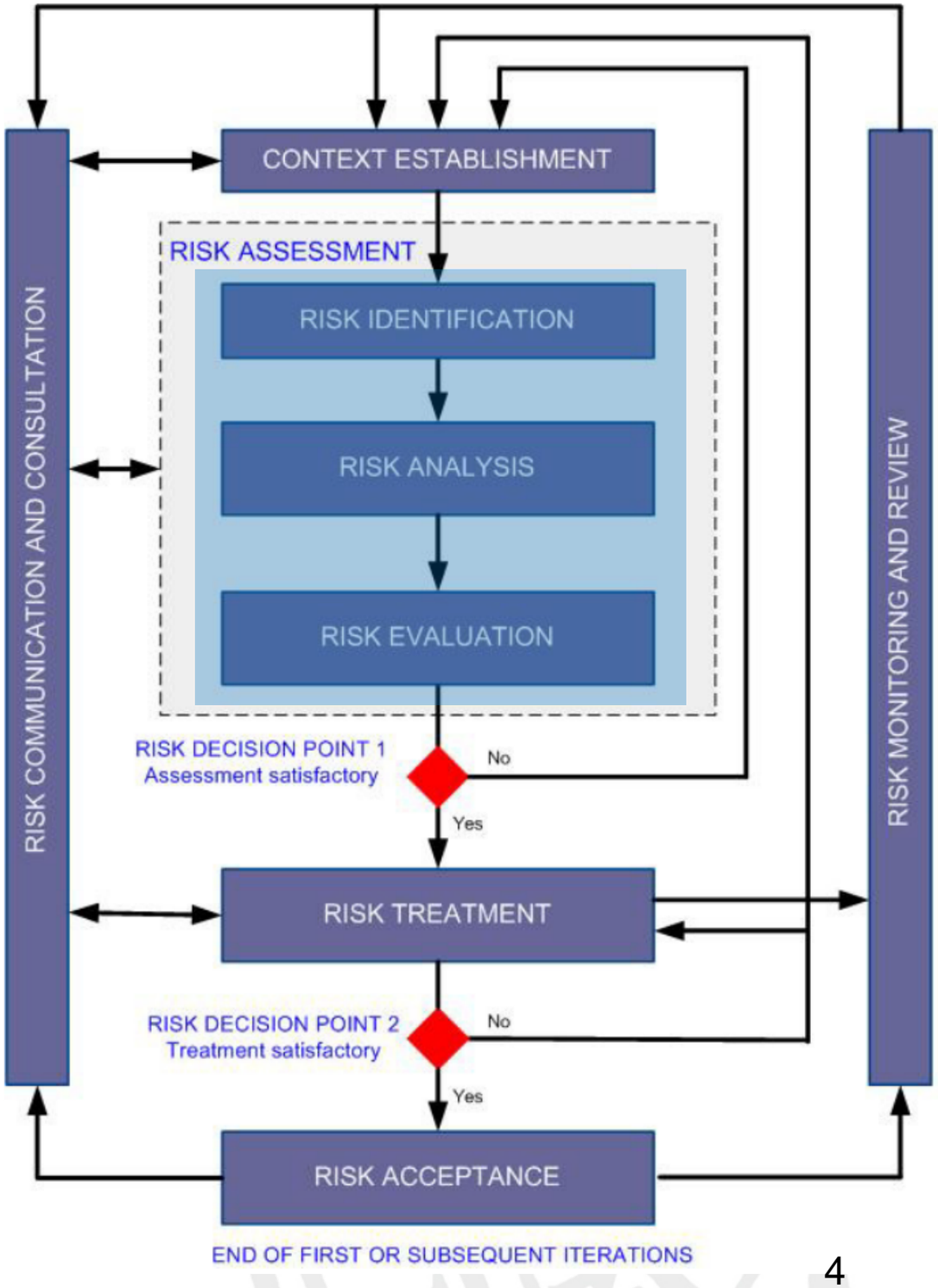
Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Objectives

- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment



[ISO 31000]

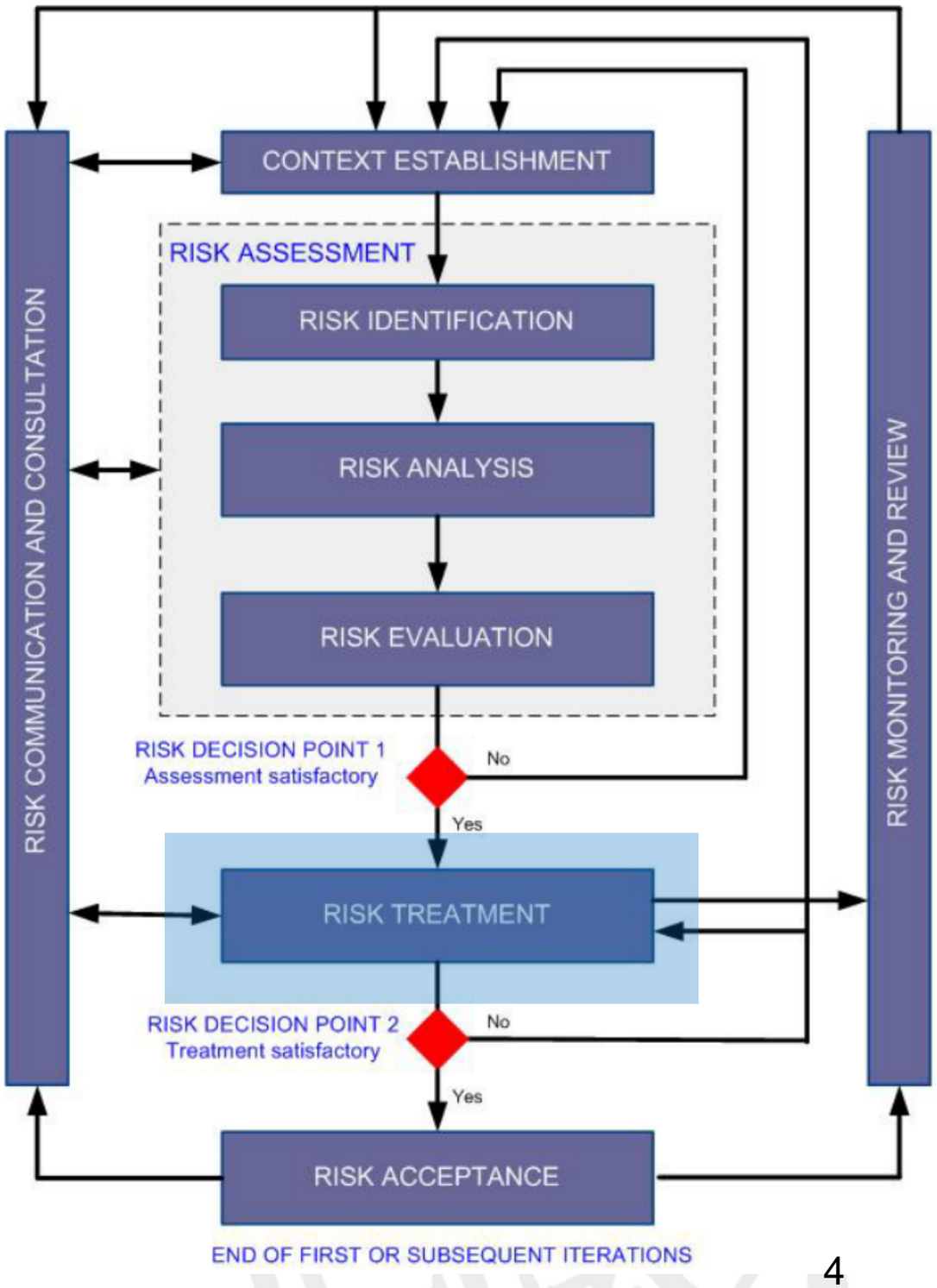
Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

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



[ISO 31000]

Risk Management vs Control System Design

Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Objectives

- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment

- Risk Mitigation
  - Prevention, Detection, Treatment

Risk Management vs Control System Design

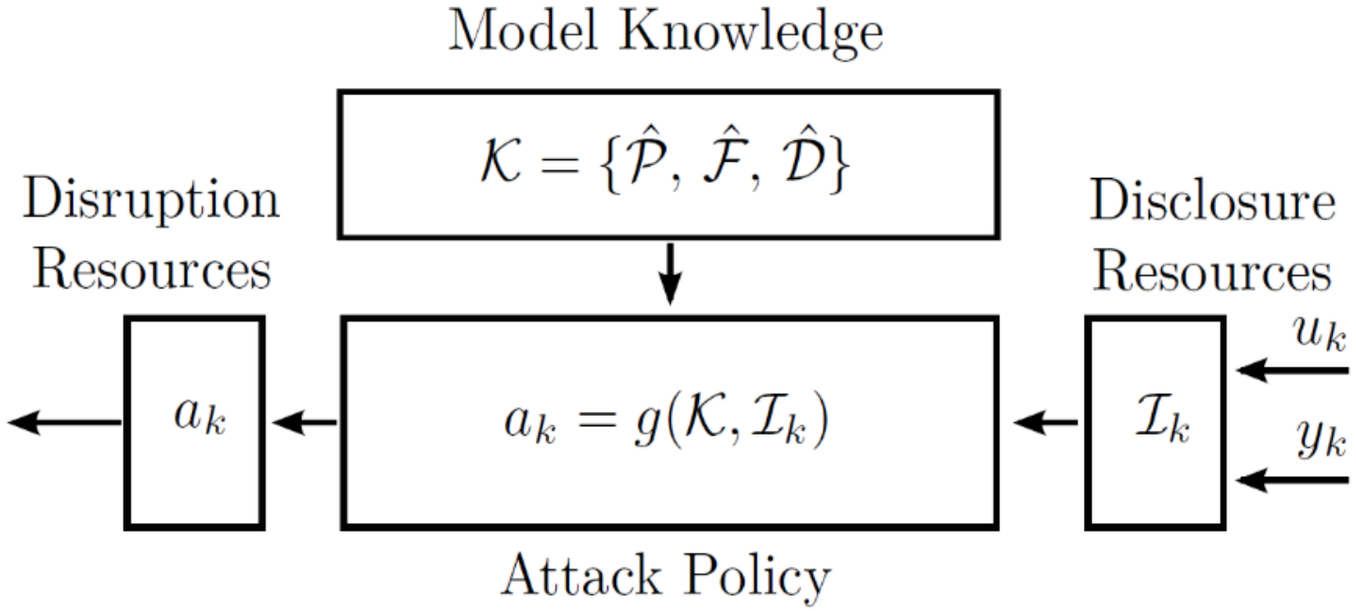
Main steps in risk management

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

Main steps in control system design:

- Scenario characterization
  - Models, Scenarios, Cost functions
- System Analysis
  - Structural: Controllability, observ. …
  - Performance: Robustness, optimality
- System Design
  - Structure of the system
  - Monitoring algorithms
  - Control algorithms

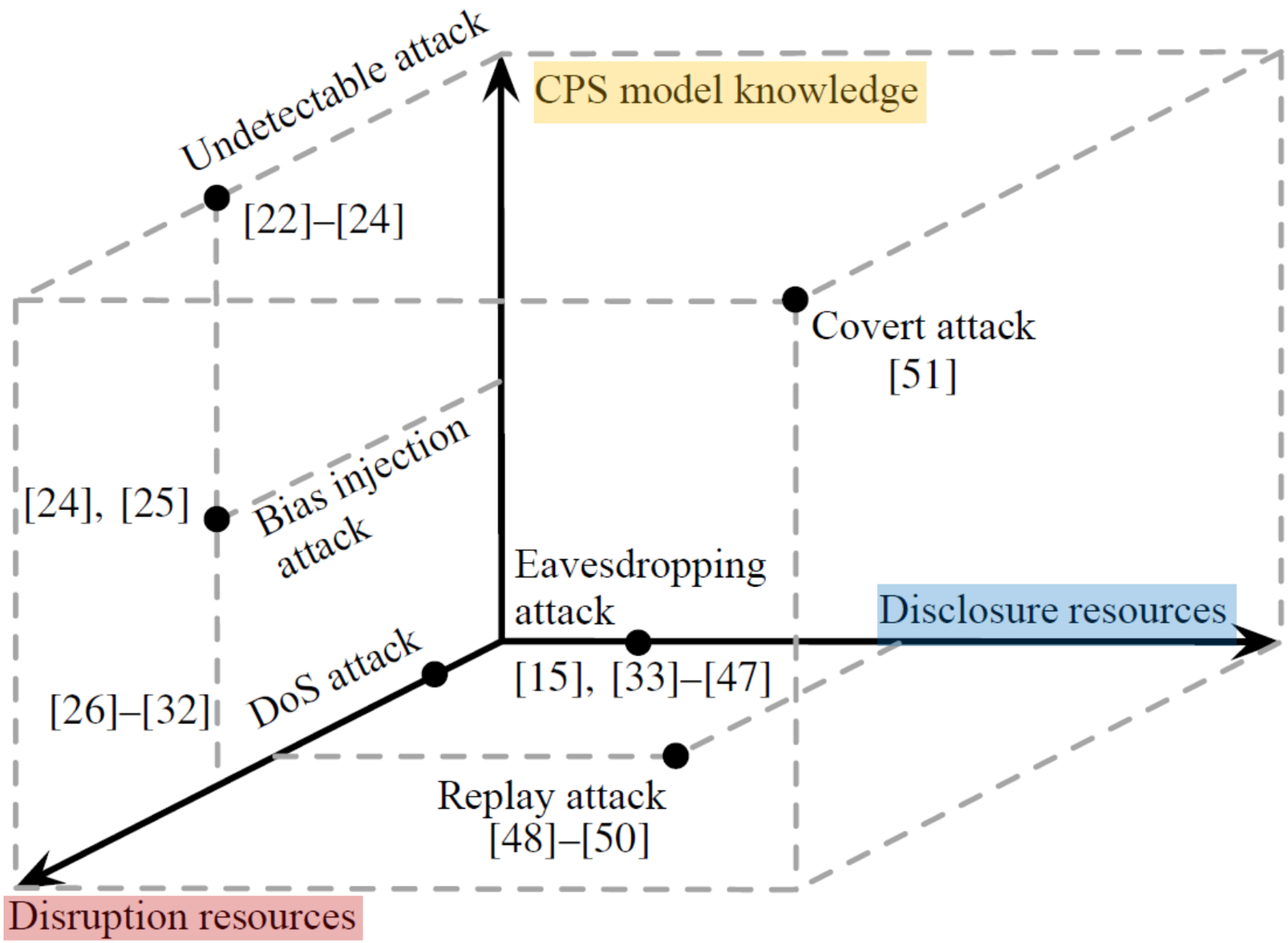
Overview - Adversary Models

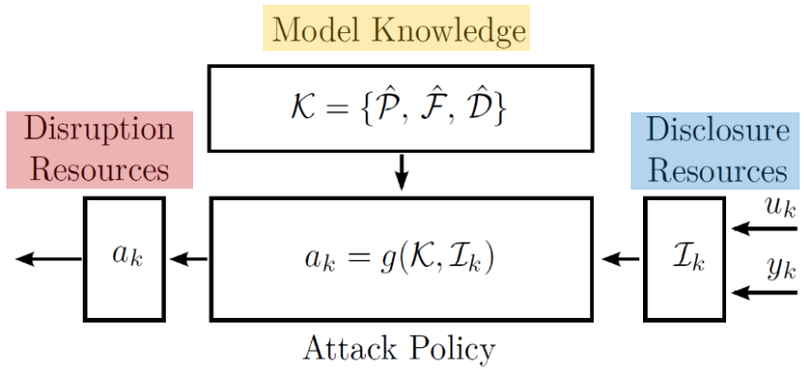


[Teixeira *et al*., Automatica, 2015]

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

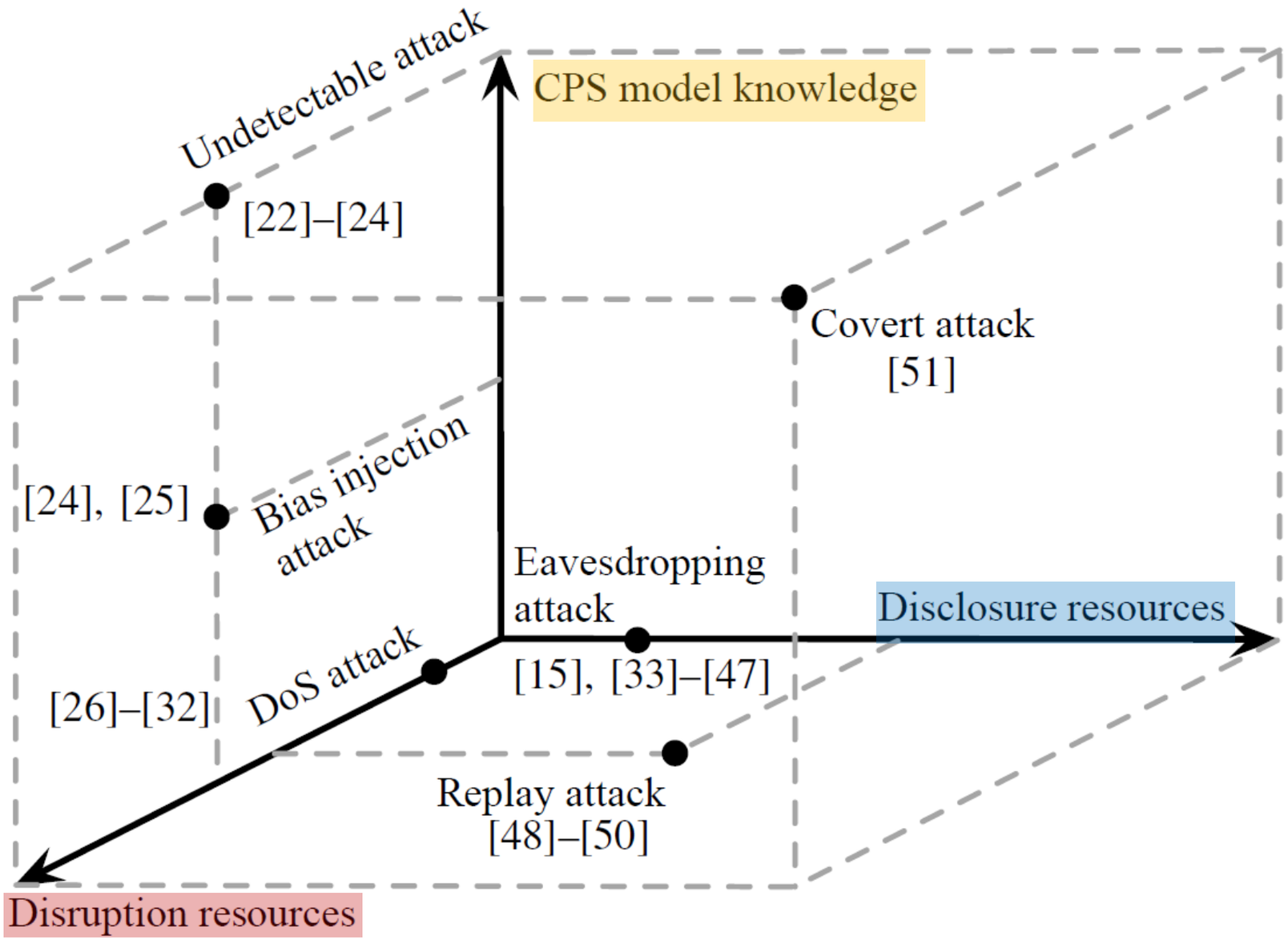
Overview - Attack Scenarios

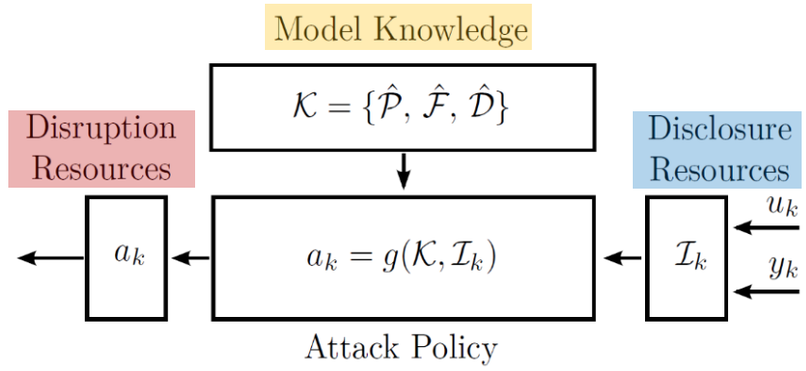




[Chong *et al.*, ECC, 2019] [Teixeira *et al.*, Automatica, 2015]

Overview - Attack Scenarios

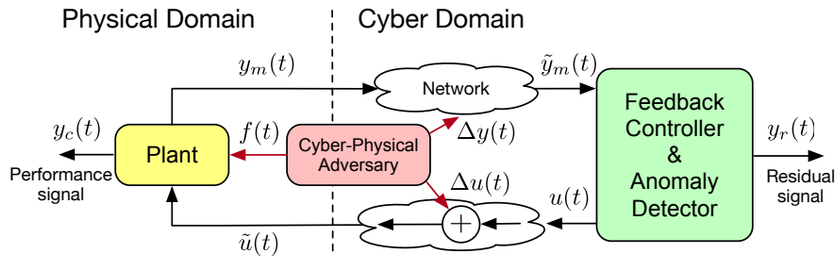




- **Adversary Models**
  - How does the adversary behave?
- **Security Analysis:**
  - Can it evade detection?
  - Can it violate safety?
- **Tailored Detection / Mitigation Schemes**

[Chong *et al*., ECC, 2019] [Teixeira *et al*., Automatica, 2015]

Undetectable Attacks



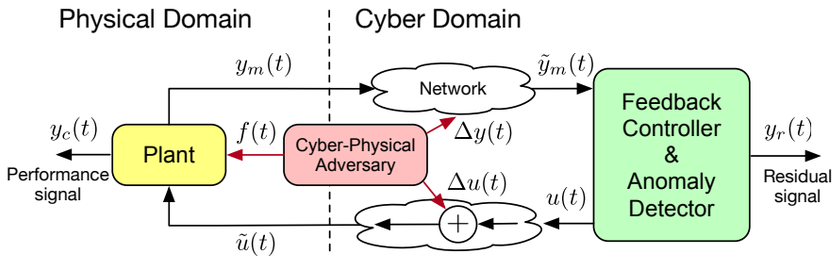
System dynamics:

$$x(k+1) = Ax(k) + Bu(k) + B_a a(k)$$

$$y(k) = Cx(k) + D_a a(k)$$

$$a_k = \begin{bmatrix} \Delta u_k \\ \Delta y_k \end{bmatrix}$$

Undetectable Attacks



System dynamics:

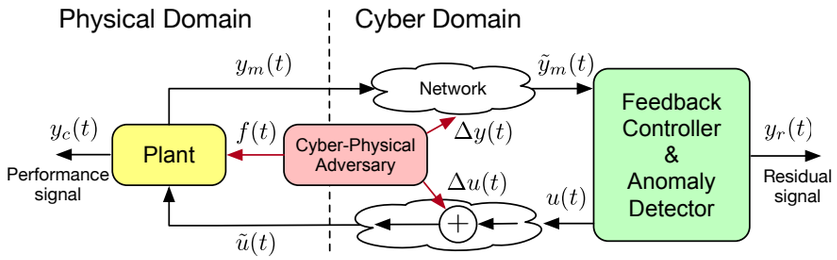
$$x(k+1) = Ax(k) + Bu(k) + B_a a(k)$$

$$y(k) = Cx(k) + D_a a(k)$$

$$a_k = \begin{bmatrix} \Delta u_k \\ \Delta y_k \end{bmatrix}$$

Output function: $y(k) = \Phi(x_0, a, k) \triangleq CA^k x_0 + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$

Undetectable Attacks



System dynamics:

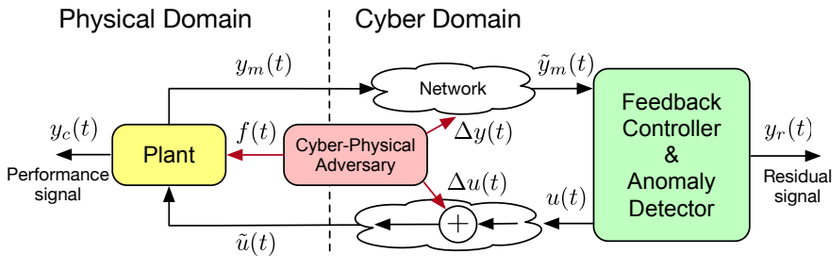
$$x(k+1) = Ax(k) + Bu(k) + B_a a(k)$$
$$y(k) = Cx(k) + D_a a(k)$$

$$a_k = \begin{bmatrix} \Delta u_k \\ \Delta y_k \end{bmatrix}$$

Output function: $y(k) = \Phi(x_0, a, k) \triangleq CA^k x_0 + C\sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$

Measurement trajectory under attack a: $y(k) = \Phi(x_0, a, k), \quad k \geq 0$

Undetectable Attacks



System dynamics:

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

$$a_k = \begin{bmatrix} \Delta u_k \\ \Delta y_k \end{bmatrix}$$

Output function: $y(k) = \Phi(x_0, a, k) \triangleq CA^k x_0 + C\sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$

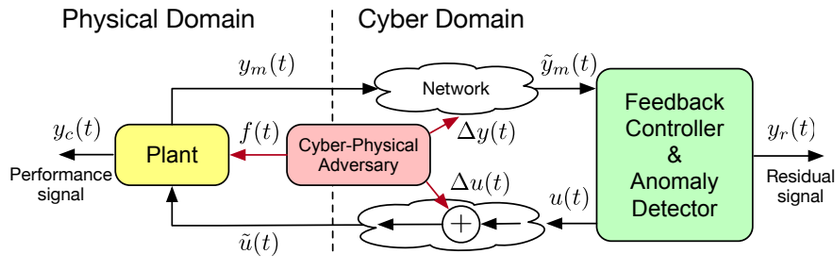
Measurement trajectory under attack a: $y(k) = \Phi(x_0, a, k), \quad k \geq 0$

Definition: Attack a is *undetectable* if

$$\Phi(x_0, a, k) = \Phi(x_0^a, 0, k)$$

for *some* initial state x_0^a for all $k \geq 0$

Undetectable Attacks



System dynamics:

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

$$a_k = \begin{bmatrix} \Delta u_k \\ \Delta y_k \end{bmatrix}$$

Output function: $y(k) = \Phi(x_0, a, k) \triangleq CA^k x_0 + C\sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$

Measurement trajectory under attack a: $y(k) = \Phi(x_0, a, k), \quad k \geq 0$

Definition: Attack a is *undetectable* if

$$\Phi(x_0, a, k) = \Phi(x_0^a, 0, k)$$

for *some* initial state x_0^a for all $k \geq 0$

Interpretation: the output under attack can be confused as the result of an initial state without attack

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(x_0, a, k) - \Phi(x_0^a, 0, k)$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(x_0, a, k) - \Phi(x_0^a, 0, k)$$

[linearity] $$0 = \Phi(x_0 - x_0^a, a, k)$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(x_0, a, k) - \Phi(x_0^a, 0, k)$$

[linearity] $$0 = \Phi(x_0 - x_0^a, a, k)$$

$$0 = CA^k \underbrace{(x_0 - x_0^a)}_{\triangleq \bar{x}_0^a} + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(x_0, a, k) - \Phi(x_0^a, 0, k)$$

[linearity] $$0 = \Phi(x_0 - x_0^a, a, k)$$

$$0 = CA^k \underbrace{(x_0 - x_0^a)}_{\triangleq \bar{x}_0^a} + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

$$0 = \Phi(\bar{x}_0^a, a, k)$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(x_0, a, k) - \Phi(x_0^a, 0, k)$$

[linearity] $$0 = \Phi(x_0 - x_0^a, a, k)$$

$$0 = CA^k \underbrace{(x_0 - x_0^a)}_{\triangleq \bar{x}_0^a} + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

$$0 = \Phi(\bar{x}_0^a, a, k)$$

$\Leftrightarrow$ Must exist initial state $\bar{x}_0^a$ and input a_k yielding zero output.

This corresponds to the zero dynamics of the system: a well-known concept in control theory!

Undetectable Attacks and the Zero Dynamics

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(\bar{x}_0^a, a, k) = CA^k \bar{x}_0^a + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(\bar{x}_0^a, a, k) = CA^k \bar{x}_0^a + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

Theorem: There exists undetectable attacks if and only if exist $\nu \in \mathbb{C}$ and $\bar{x}_0^a \in \mathbb{C}^{n_x}$, $g \in \mathbb{C}^{n_a}$ such that

$$\begin{bmatrix} \nu I_{n_x} - A & -B_a \\ C & D_a \end{bmatrix} \begin{bmatrix} \bar{x}_0^a \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(\bar{x}_0^a, a, k) = CA^k \bar{x}_0^a + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

Theorem: There exists undetectable attacks if and only if exist $\nu \in \mathbb{C}$ and $\bar{x}_0^a \in \mathbb{C}^{n_x}$, $g \in \mathbb{C}^{n_a}$ such that

$$\begin{bmatrix} \nu I_{n_x} - A & -B_a \\ C & D_a \end{bmatrix} \begin{bmatrix} \bar{x}_0^a \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

An undetectable attack is $a_k = \nu^k g$, which induces an impulse response with initial condition $x_0^a = x_0 - \bar{x}_0^a$.

Undetectable Attacks and the Zero Dynamics

Undetectability requires (for all $k \geq 0$)

$$0 = \Phi(\bar{x}_0^a, a, k) = CA^k \bar{x}_0^a + C \sum_{i=1}^{k-1} A^{k-i-1} B_a a_i + D_a a_k$$

Theorem: There exists undetectable attacks if and only if exist $\nu \in \mathbb{C}$ and $\bar{x}_0^a \in \mathbb{C}^{n_x}$, $g \in \mathbb{C}^{n_a}$ such that

$$\begin{bmatrix} \nu I_{n_x} - A & -B_a \\ C & D_a \end{bmatrix} \begin{bmatrix} \bar{x}_0^a \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

An undetectable attack is $a_k = \nu^k g$, which induces an impulse response with initial condition $x_0^a = x_0 - \bar{x}_0^a$.

In general requires system knowledge $\mathcal{K} = \{\mathcal{P}, \mathcal{F}, \emptyset\} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$

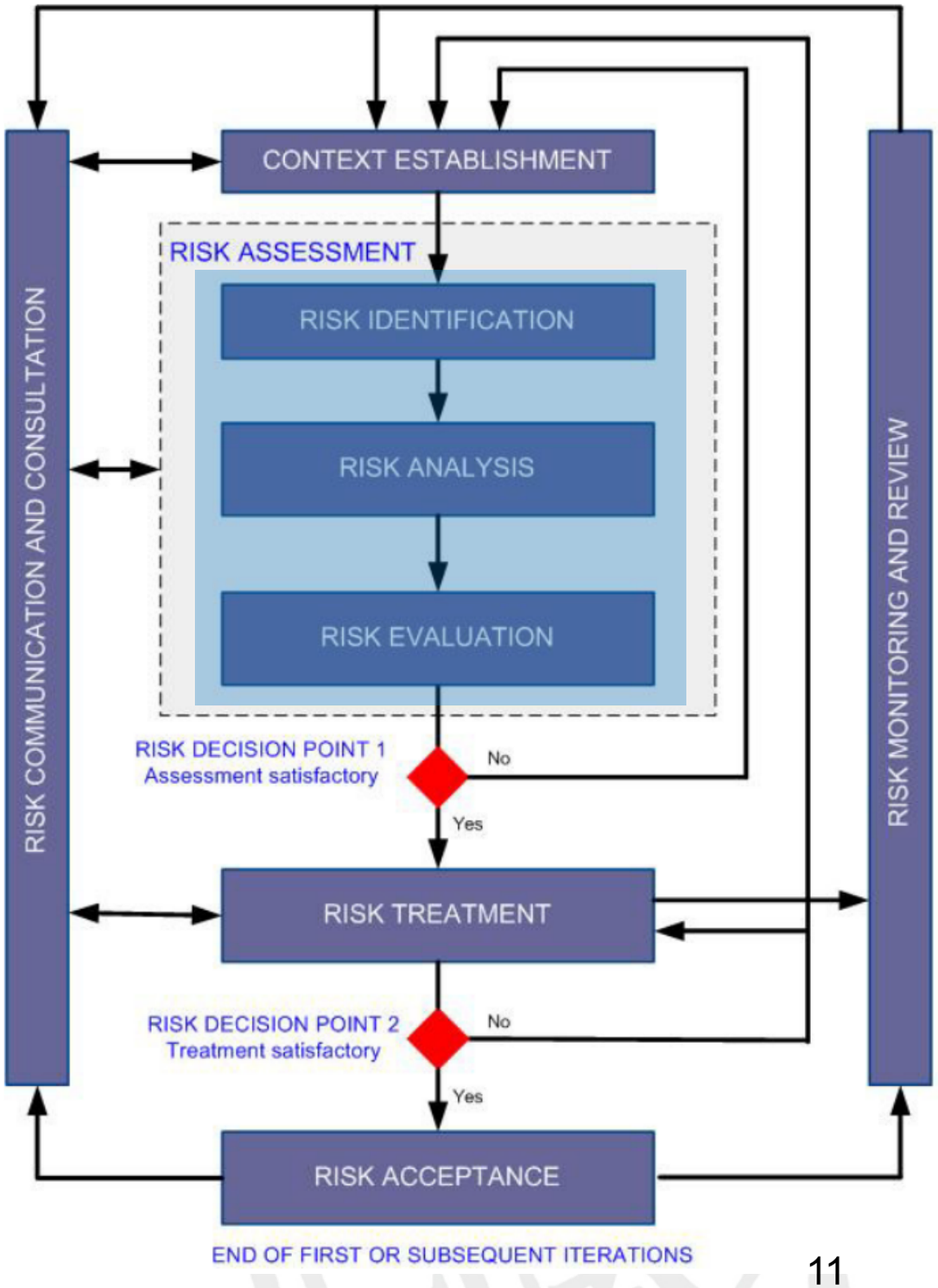
(Matlab function to compute ν: ν = tzero(ss(A, B_a, C, D_a))

Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)
[Kaplan & Garrick, 1981]

Main steps in risk management

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



[ISO 31000]

Impact analysis

Attack Model:

Impact analysis

Attack Model:

- Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Attack policy: $a_k = \nu^k g$
  - $|\nu| < 1$: vanishing attack
  - $|\nu| > 1$: diverging attack

Impact analysis

Attack Model:

- Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

- Attack policy: $a_k = \nu^k g$
  - $|\nu| < 1$: vanishing attack
  - $|\nu| > 1$: diverging attack

Impact:

A vanishing attack will decay towards zero. It has a small impact.

Impact analysis

Attack Model:

- Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

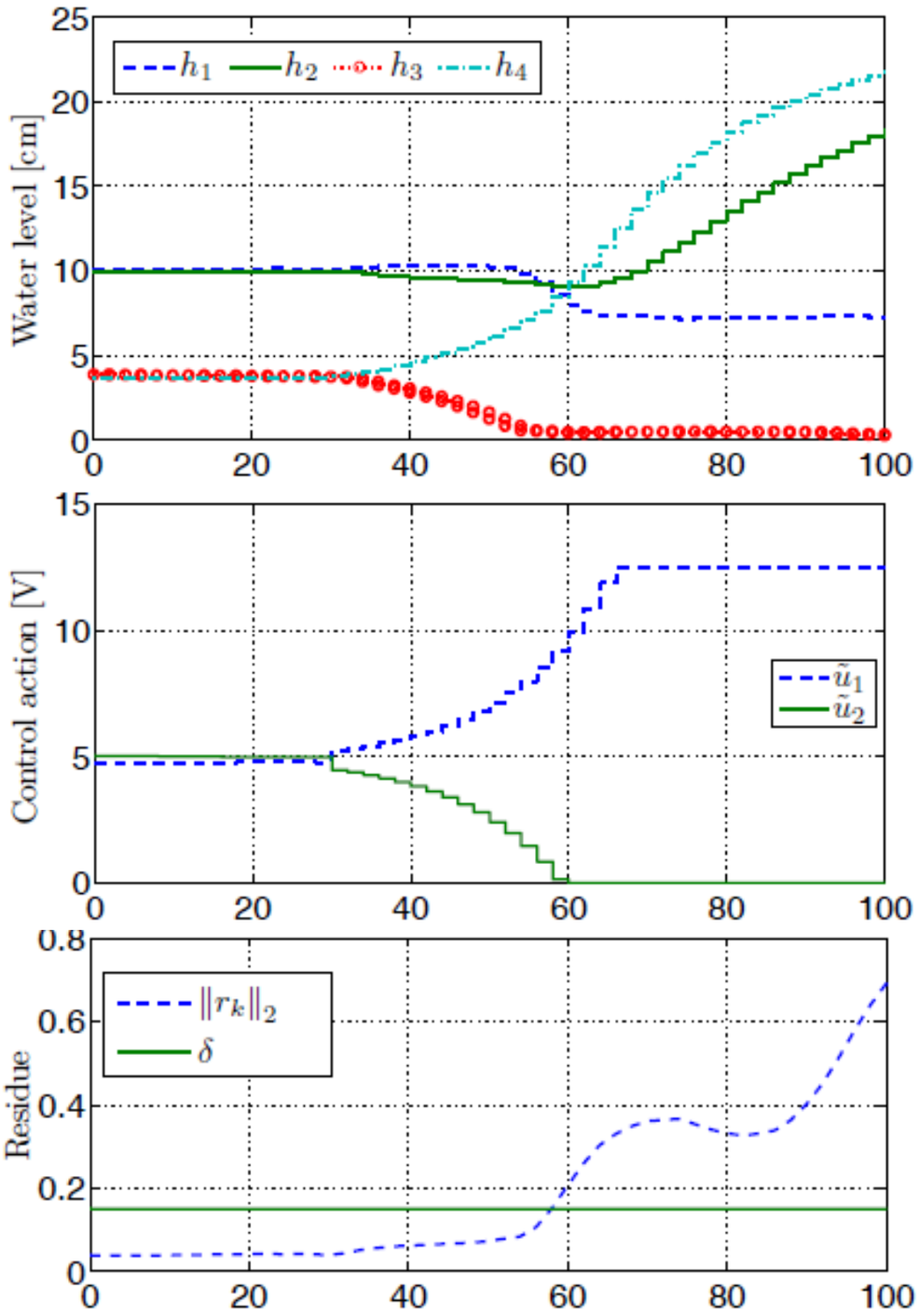
- Attack policy: $a_k = \nu^k g$
  - $|\nu| < 1$: vanishing attack
  - $|\nu| > 1$: diverging attack

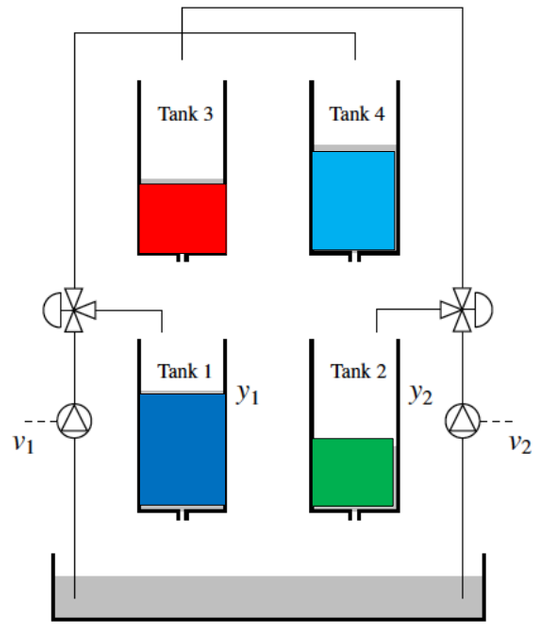
Impact:

A vanishing attack will decay towards zero. It has a small impact.

A diverging attack will increase towards infinity.
It has a large impact, if not detected.

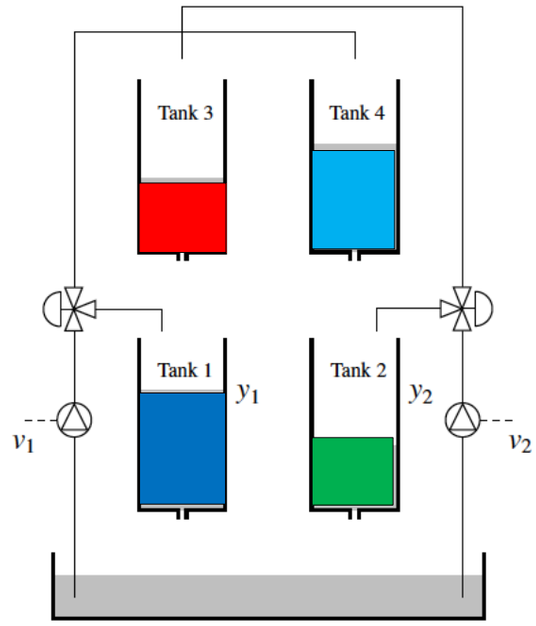
Zero Dynamics Attack - Experiment

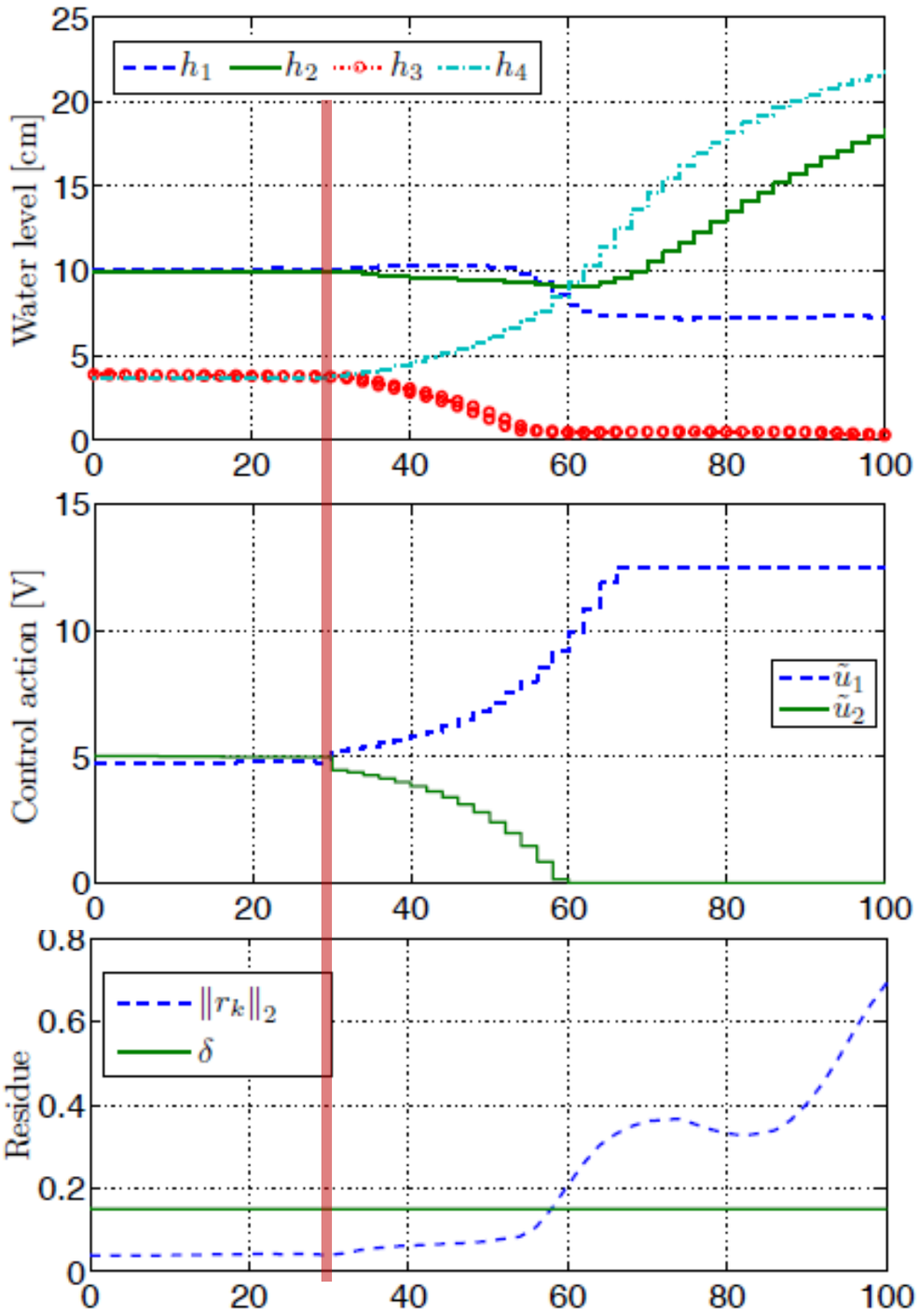




Attack Goal: Empty tank 3

Zero Dynamics Attack - Experiment

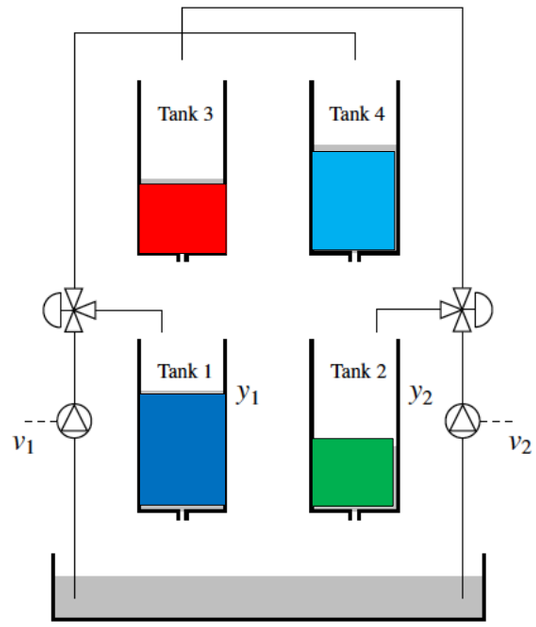


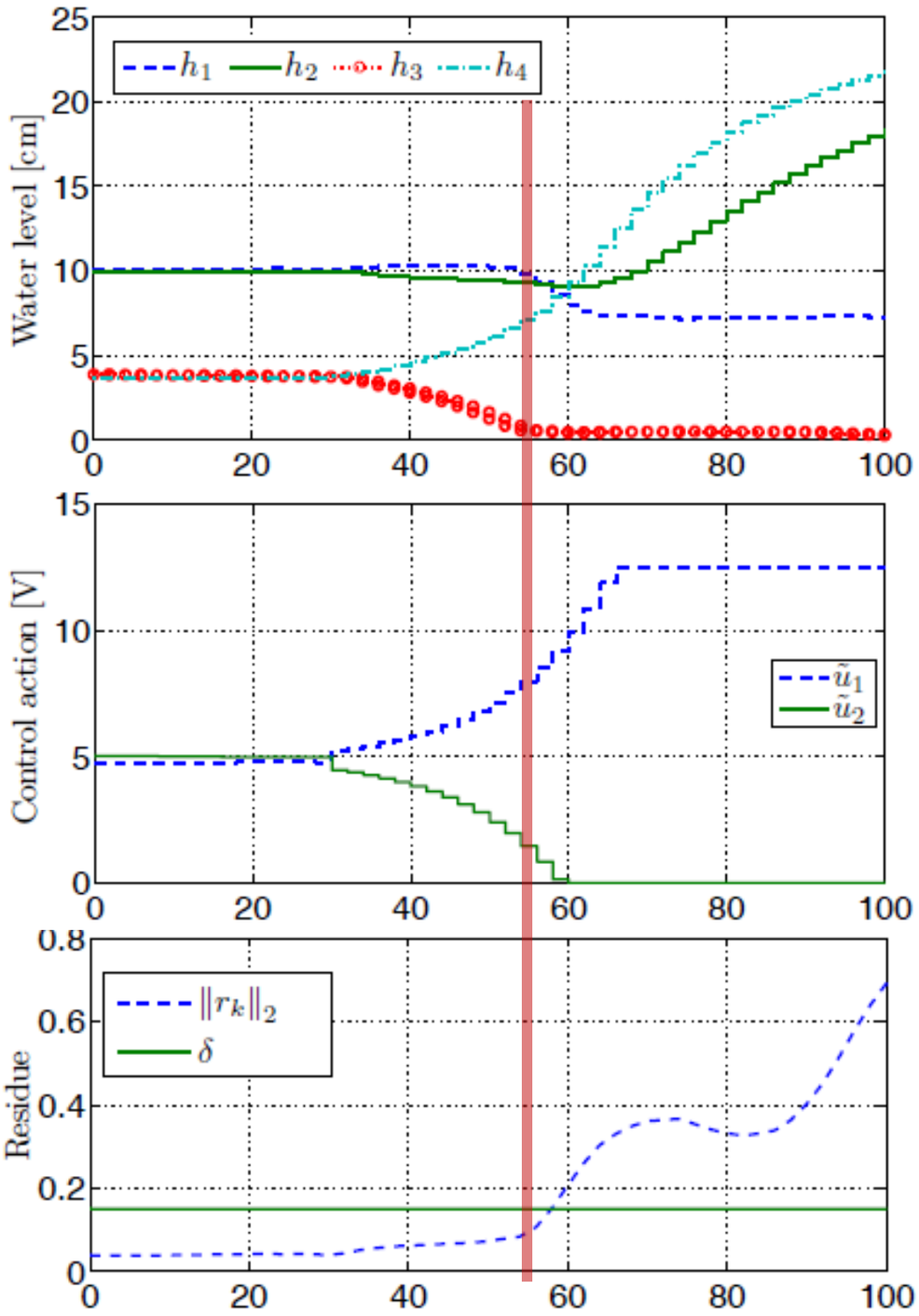


Attack Goal: Empty tank 3

- Zero dynamics attack on both actuators - unstable zero

Zero Dynamics Attack - Experiment

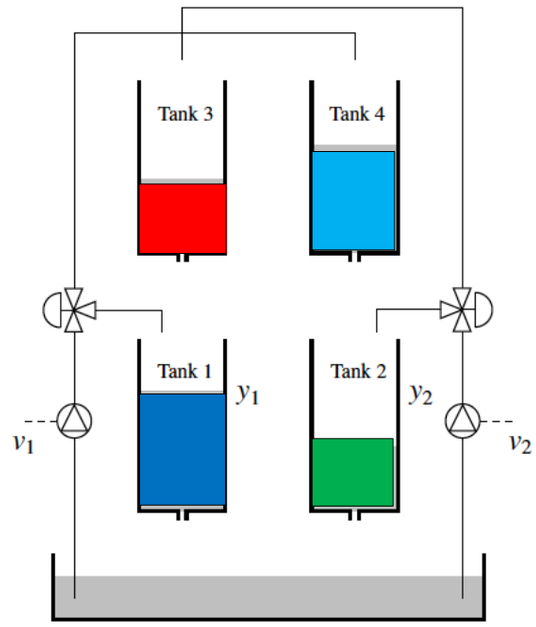


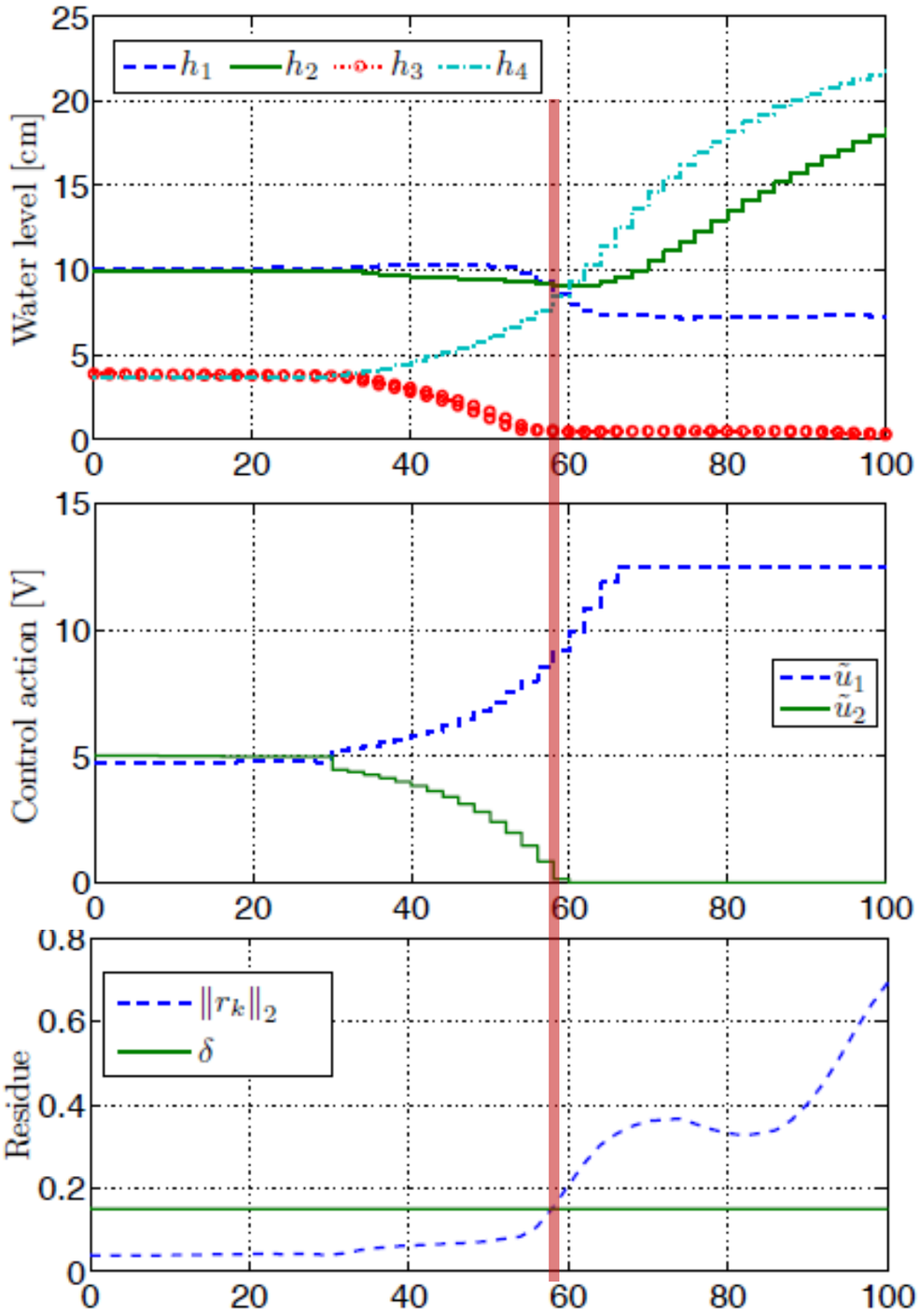


Attack Goal: Empty tank 3

- Zero dynamics attack on both actuators - unstable zero
- Tank 3 becomes empty

Zero Dynamics Attack - Experiment





Attack Goal: Empty tank 3

- Zero dynamics attack on both actuators - unstable zero
- Tank 3 becomes empty
- The attack is **detected** (Why? Can we exploit this for detection?)

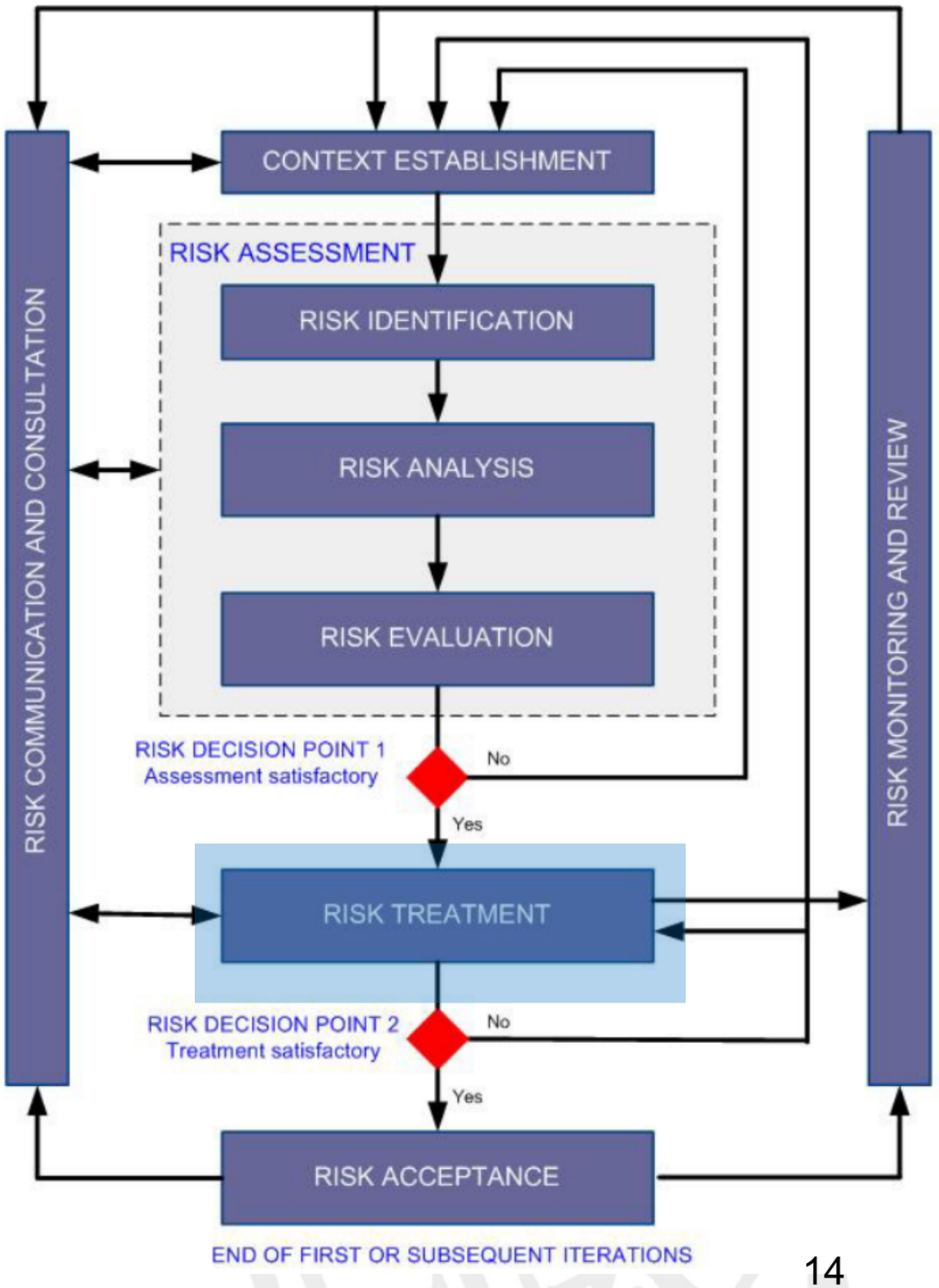
Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

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



[ISO 31000]

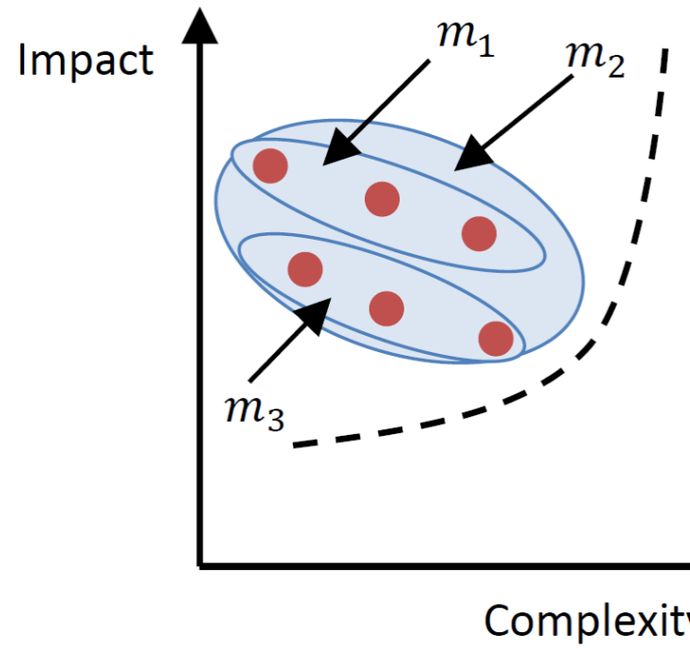
Overview - Risk Mitigation

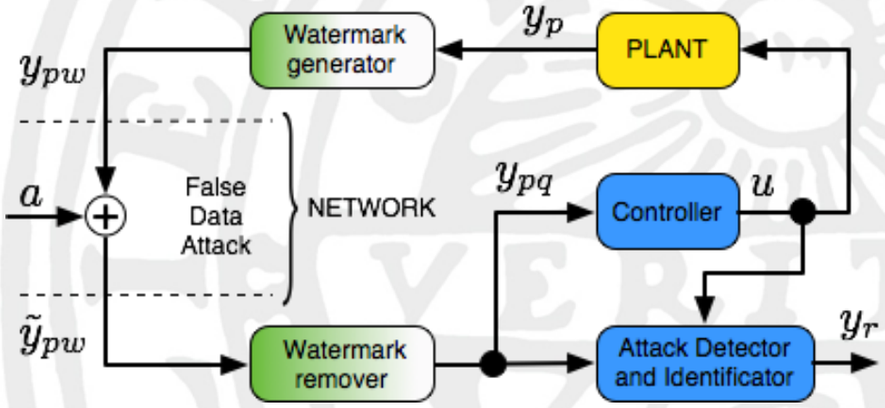
- **Prevention** (decrease likelihood by reducing vulnerability)
  - Watermarking and Moving Target Defense
  - Coding and Encryption Strategies
  - Rational Security Allocation
  - Privacy-preservation by Noise Injection

- **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
  - Resilient Control Countering DoS Attacks
  - Distributed Algorithms
  - Methods Related to Robust Statistics

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





[Ferrari and Teixeira, TAC 2021]

Control Theoretic View

Control Theoretic View

Addressing undetectable attacks:

Control Theoretic View

Addressing undetectable attacks:

- Undetectability conditions

Control Theoretic View

Addressing undetectable attacks:

- Undetectability conditions

- Binary impact assessment:
  - Is the attack impact arbitrarily large while undetected?
  - If yes, propose tailored defense (prevent or detect)

Control Theoretic View

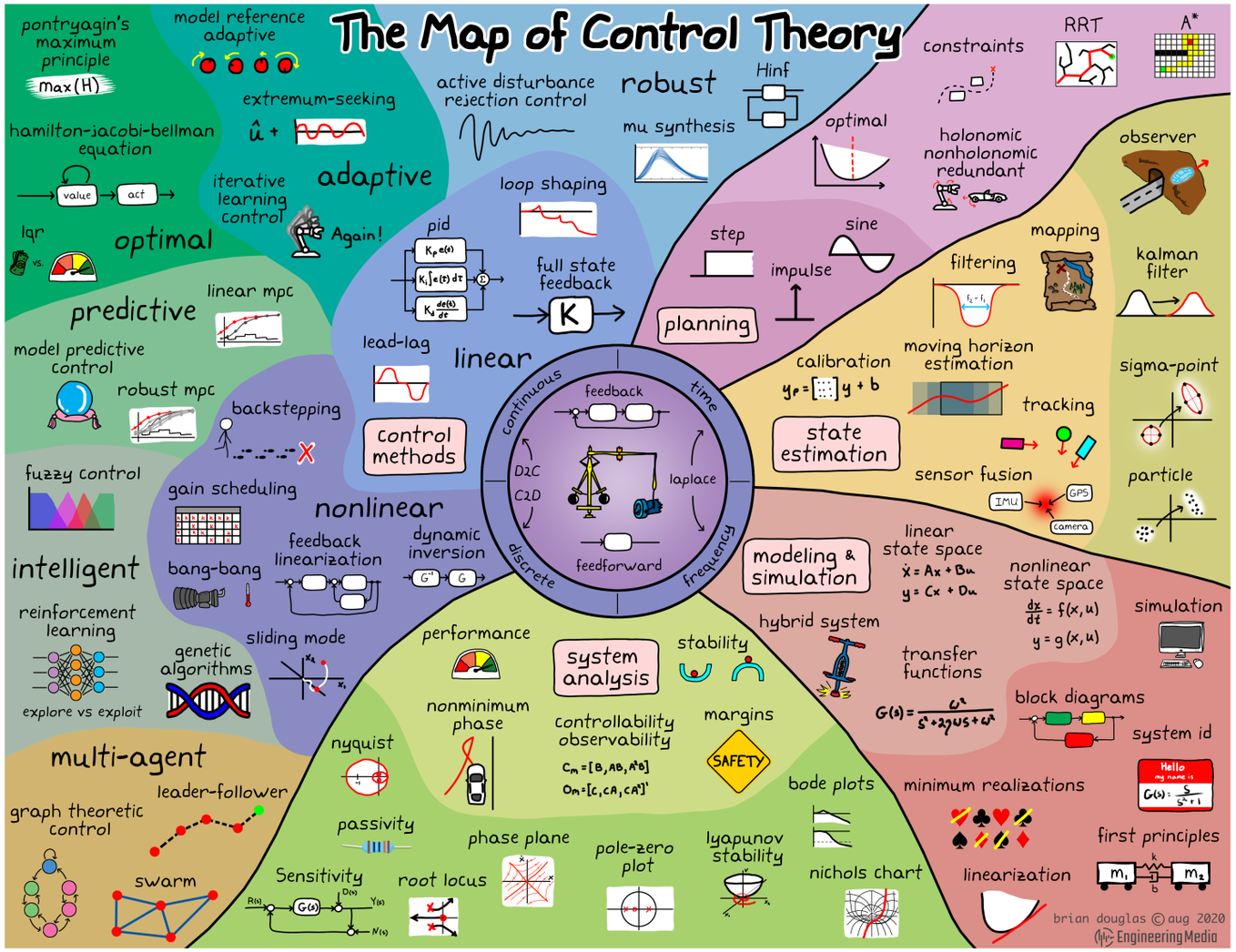
Addressing undetectable attacks:

- Undetectability conditions

- Binary impact assessment:
  - Is the attack impact arbitrarily large while undetected?
  - If yes, propose tailored defense (prevent or detect)

- Analogous to controllability analysis:
  - Is the system uncontrollable?
  - If yes, change the system (add more actuators)

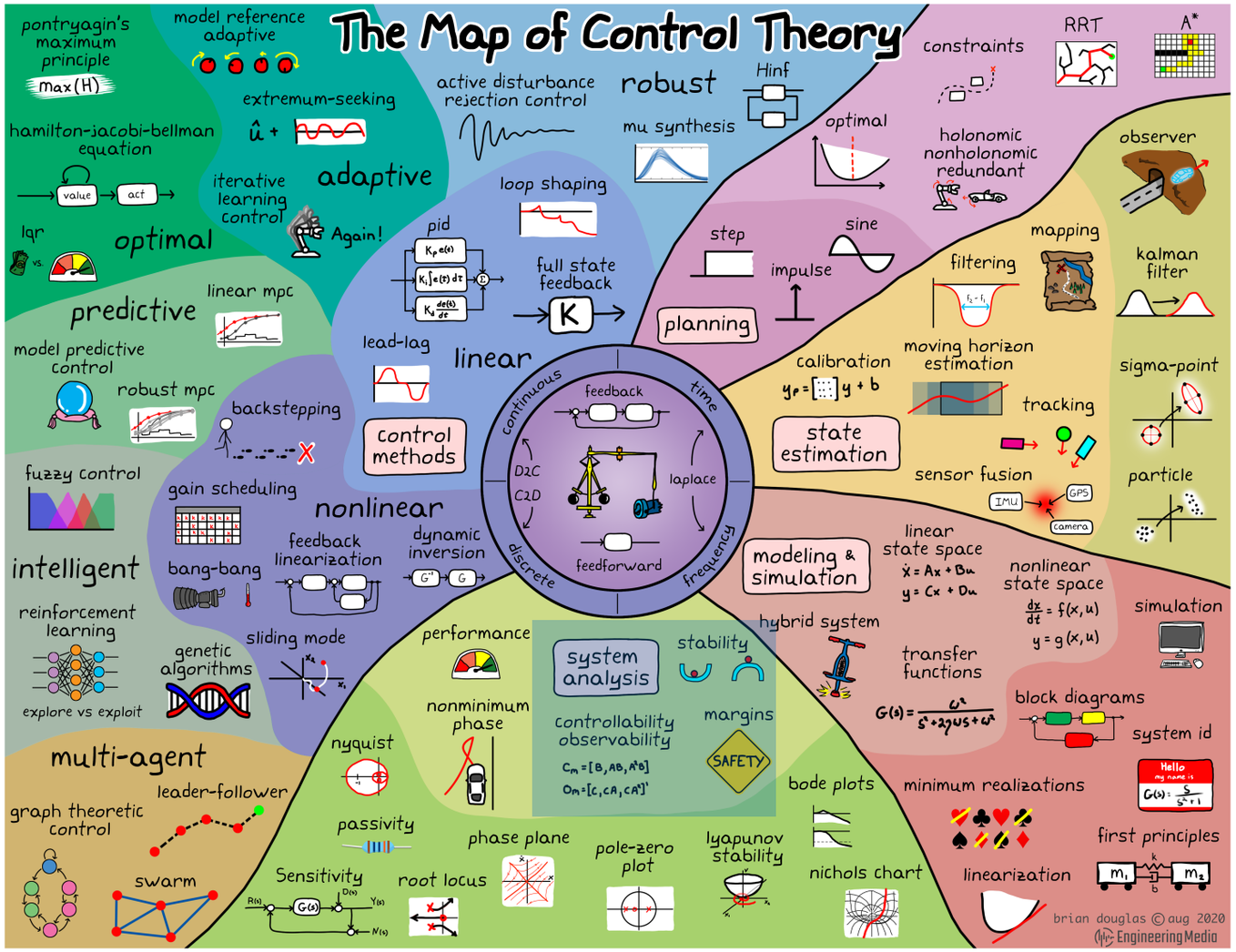
Control Theoretic View



Addressing undetectable attacks:

- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

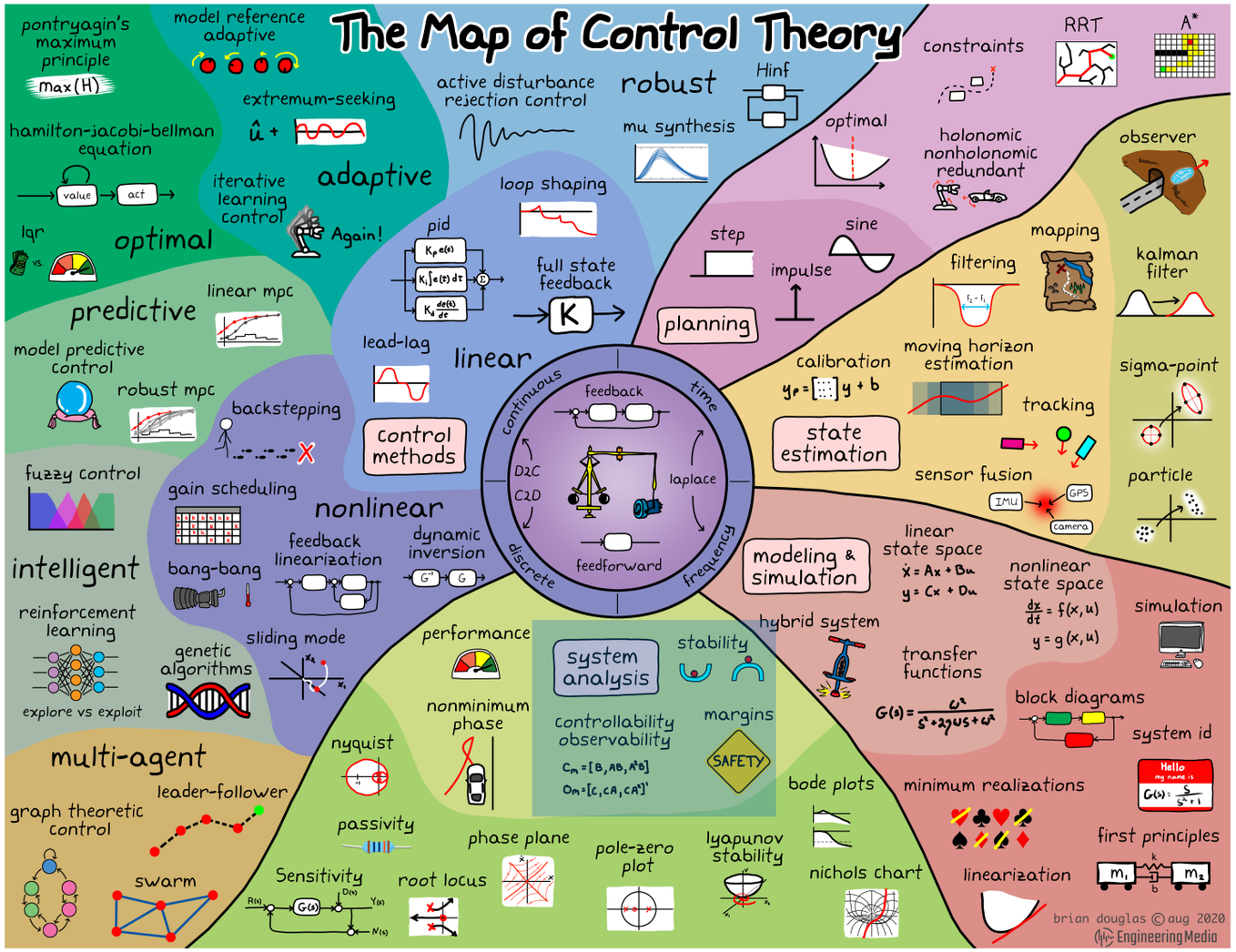
Control Theoretic View



Addressing undetectable attacks:

- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

Control Theoretic View

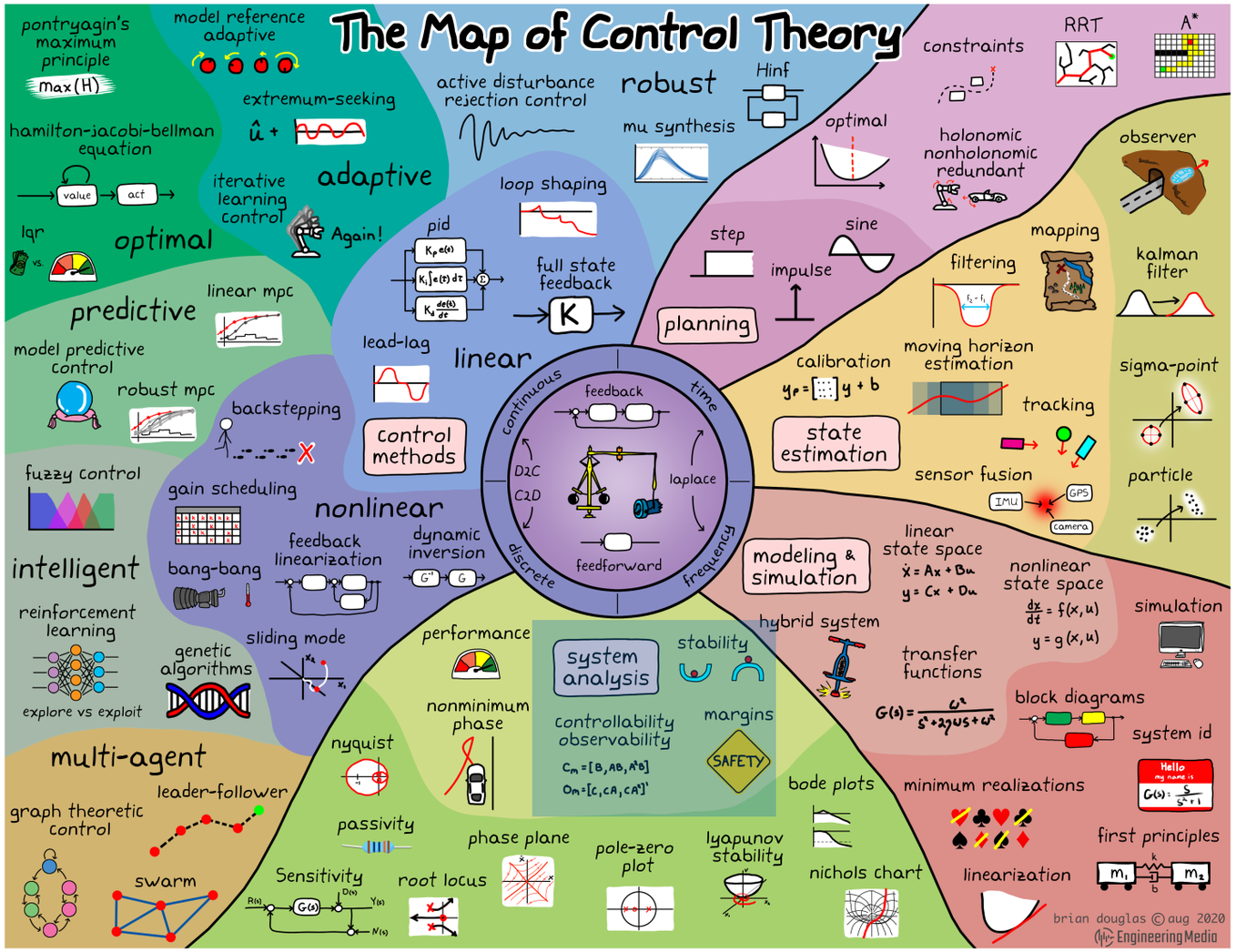


Challenges:

Addressing undetectable attacks:

- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

Control Theoretic View



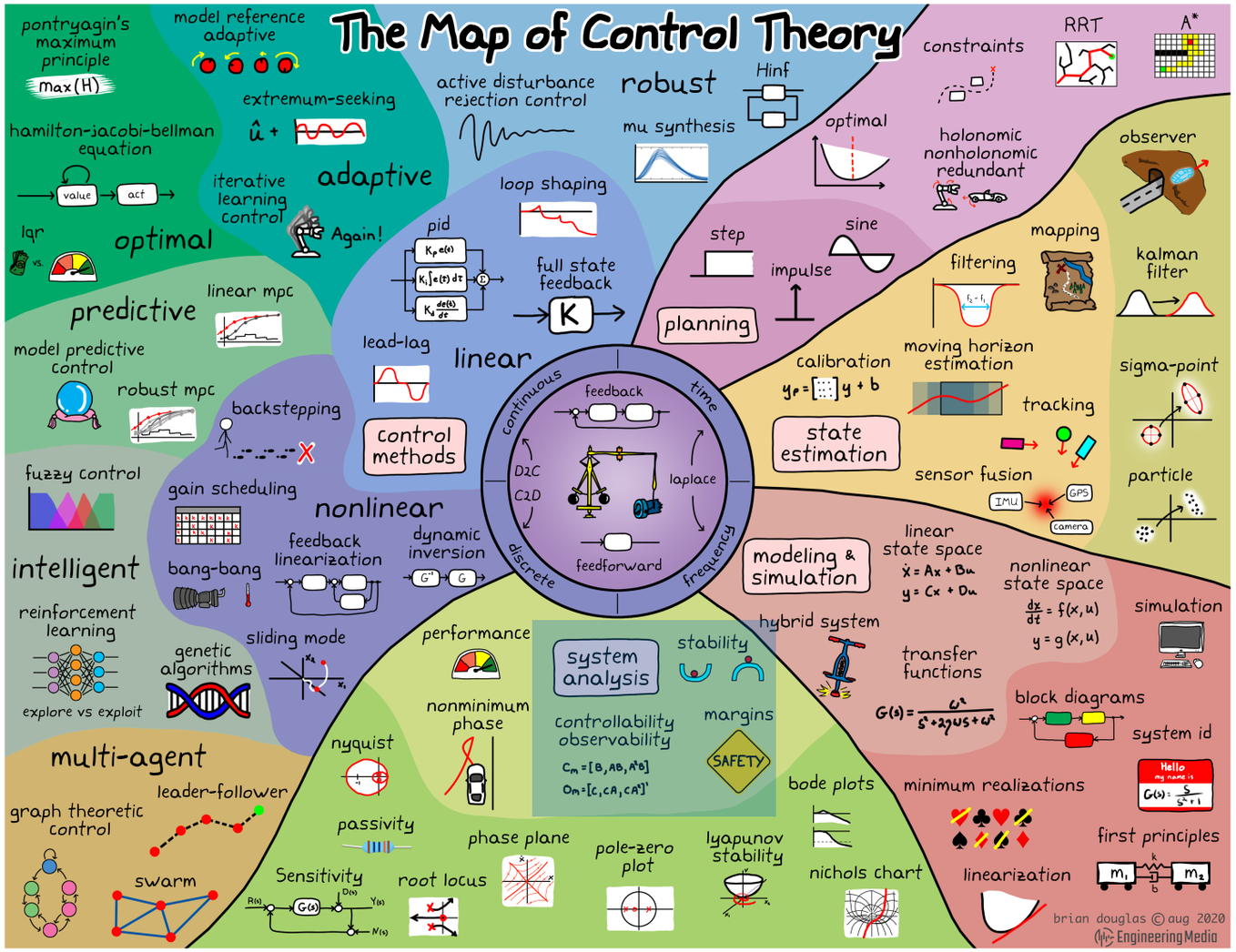
Challenges:

1. What about “weakly detectable” attacks?

Addressing undetectable attacks:

- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

Control Theoretic View



Addressing undetectable attacks:

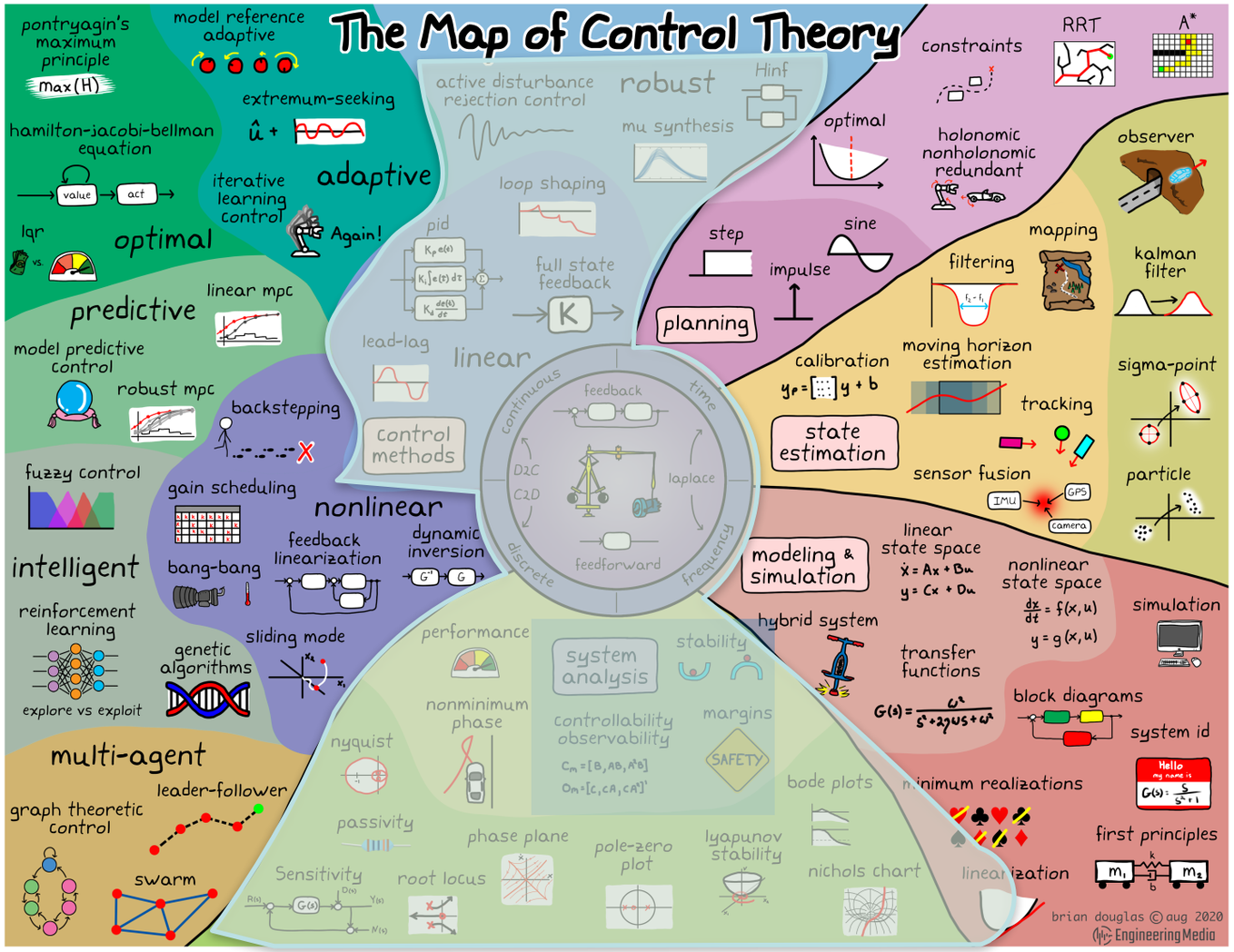
- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

Challenges:

1. What about “weakly detectable” attacks?

2. Bring control-theoretic design methods into the secure control problem

Control Theoretic View



Addressing undetectable attacks:

- Undetectability conditions
- Binary impact assessment:
  - Is the attack impact <u>arbitrarily large while undetected</u>?
  - <u>If yes</u>, propose tailored defense (prevent or detect)
- Analogous to controllability analysis:
  - Is the system <u>uncontrollable</u>?
  - <u>If yes</u>, change the system (add more actuators)

Challenges:

1. What about “weakly detectable” attacks?

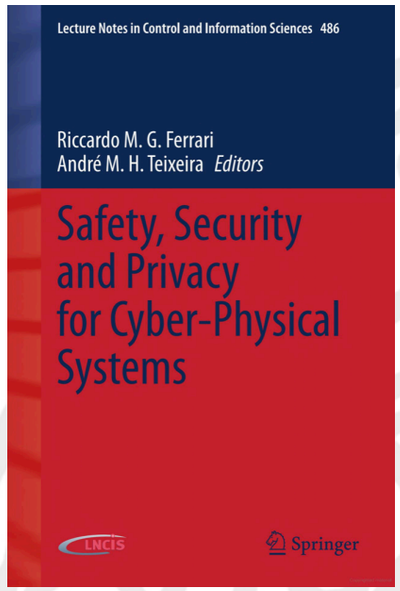
2. Bring control-theoretic design methods into the secure control problem

Outline

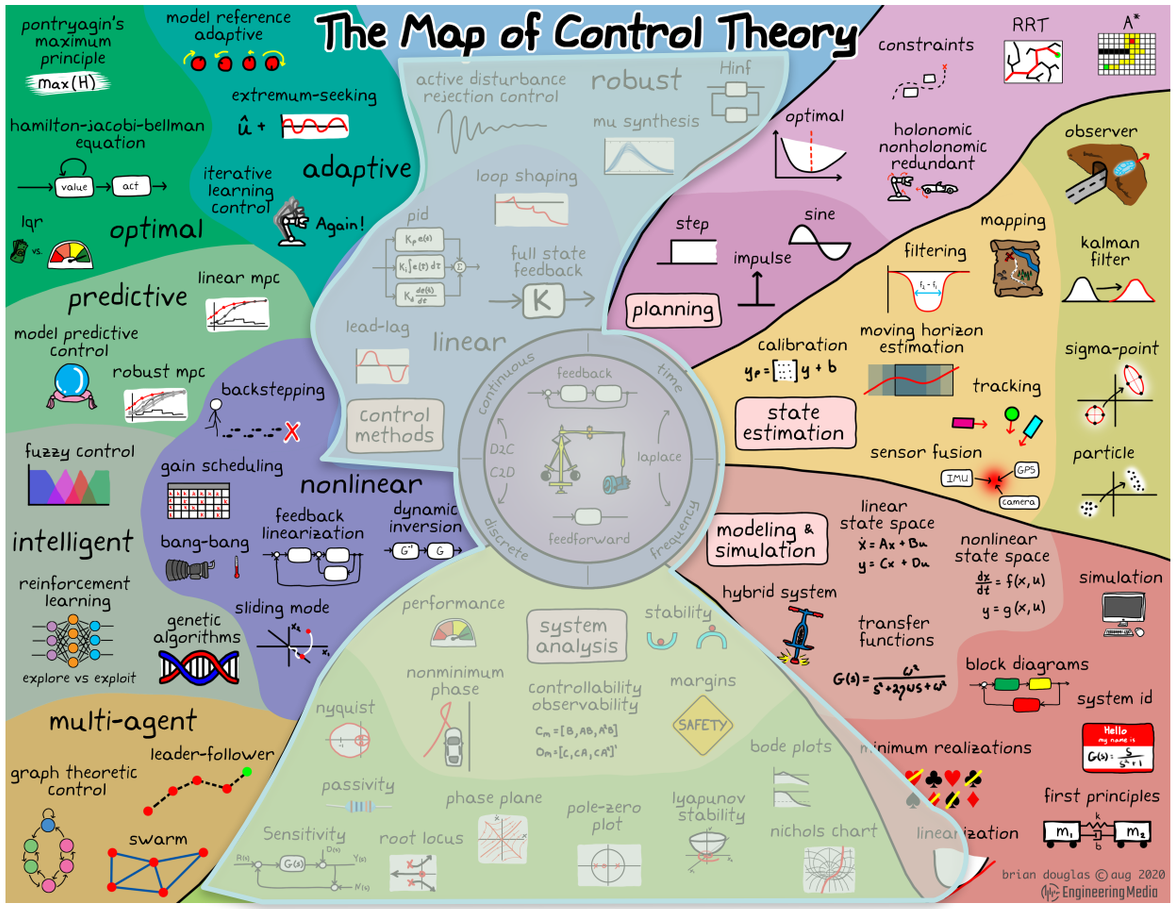
- Cybersecurity in Control Systems
  - Security Game
  - Risk Management
    - Mapping to Control System Design
    - Example: undetectable attacks

- Security Metrics for Control Systems
  - Classical metrics in control engineering
  - Novel security metric for analysis and design
    - Analysis and design problems
    - Structural limitations - invariant zeros

- Additional topics
  - Variations of the adversary models and their respective metrics
  - Incorporating uncertainty for robust open-loop attacks



Control System Design



Main steps in control system design:

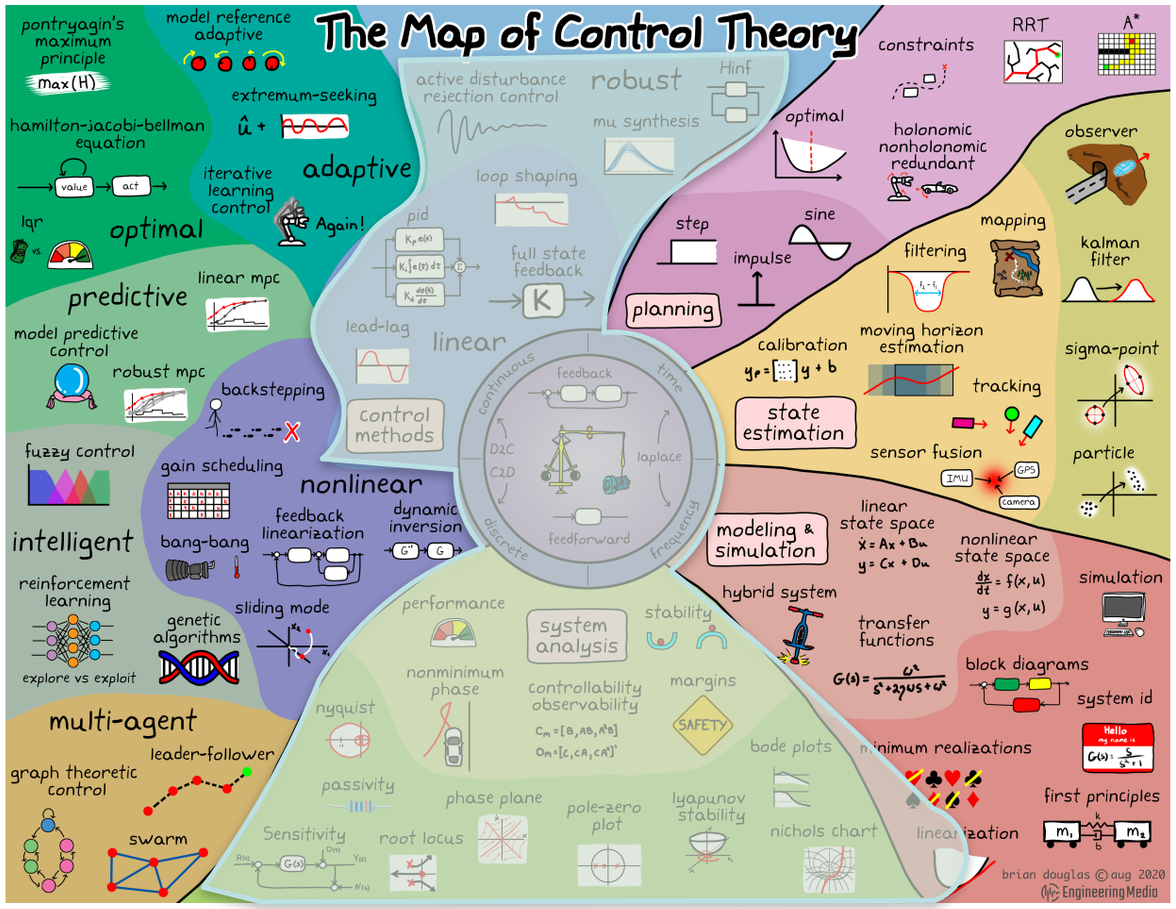
- Scenario characterization
  - Models, Scenarios, Cost functions

- System Analysis
  - Sensitivity metrics (w.r.t. input) [not binary in general]

- System Design (Control Methods)
  - Find control and monitoring algorithms that minimise the sensitivity metric

  - Re-design the structure of the system to minimise the sensitivity metric

Control System Design



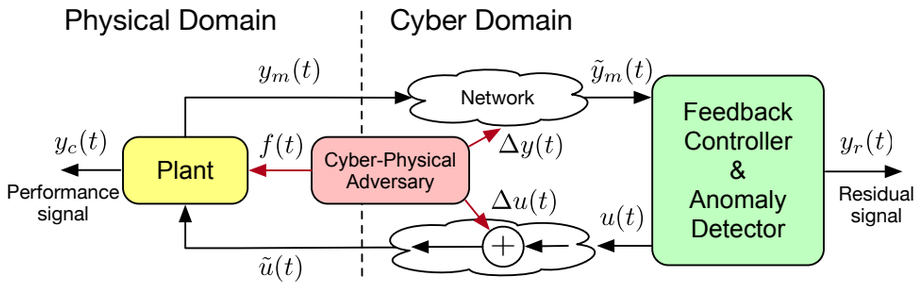
Main steps in control system design:

- Scenario characterization
  - Models, Scenarios, Cost functions

- System Analysis
  - Sensitivity metrics (w.r.t. input) [not binary in general]

- System Design (Control Methods)
  - Find control and monitoring algorithms that minimise the sensitivity metric

  - Re-design the structure of the system to minimise the sensitivity metric



Adversary Model

$$y_p(t)$$

$$\dot{x}(t) = Ax(t) + Ba(t)$$
$$y_p(t) = C_p x(t) + D_p a(t)$$
$$y_r(t) = C_r x(t) + D_r a(t)$$

$a(t)$

$$y_r(t)$$

Closed-loop system

Model knowledge

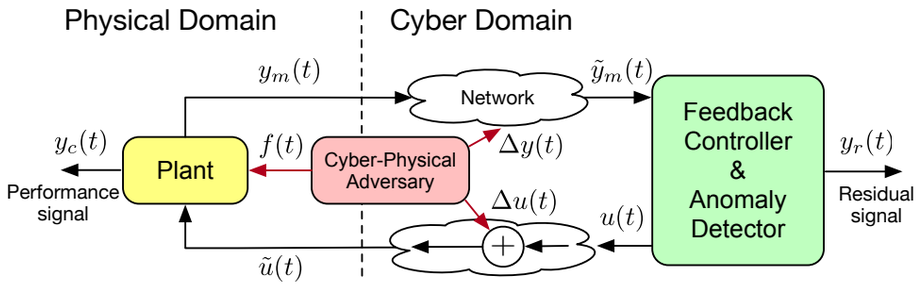
$$\{\Sigma_p,\ \Sigma_r\}$$

Disruption Resources

$a(t)$

Attack Policy

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



Adversary Model

$$y_p(t)$$

$$\dot{x}(t) = Ax(t) + Ba(t)$$
$$y_p(t) = C_p x(t) + D_p a(t)$$
$$y_r(t) = C_r x(t) + D_r a(t)$$

$$a(t)$$

$$y_r(t)$$

Closed-loop system

Model knowledge

$$\{\Sigma_p,\ \Sigma_r\}$$

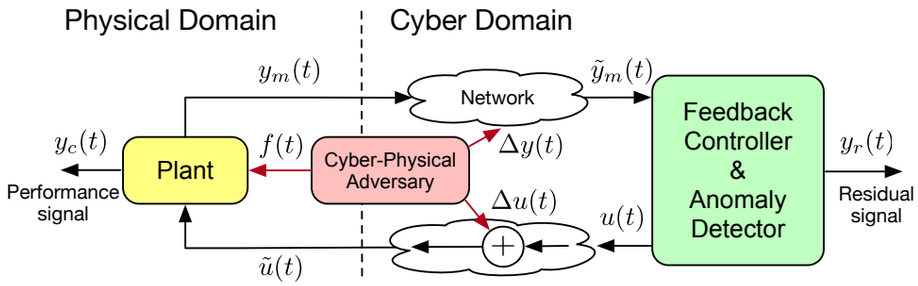
Disruption Resources

$$a(t)$$

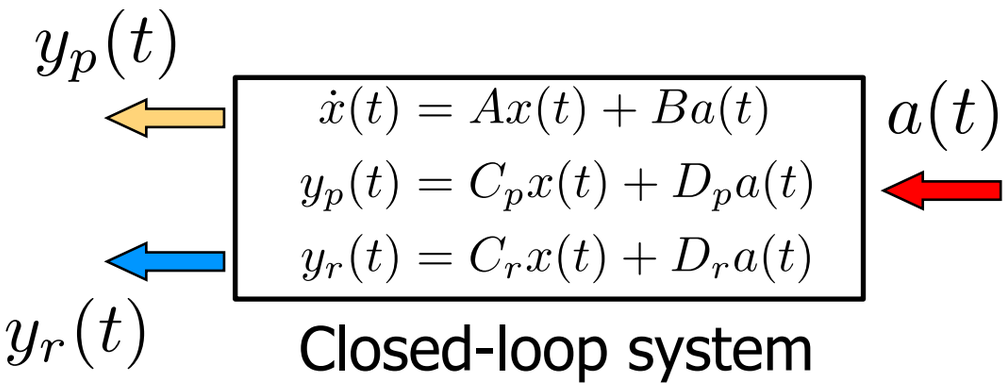
Attack Policy

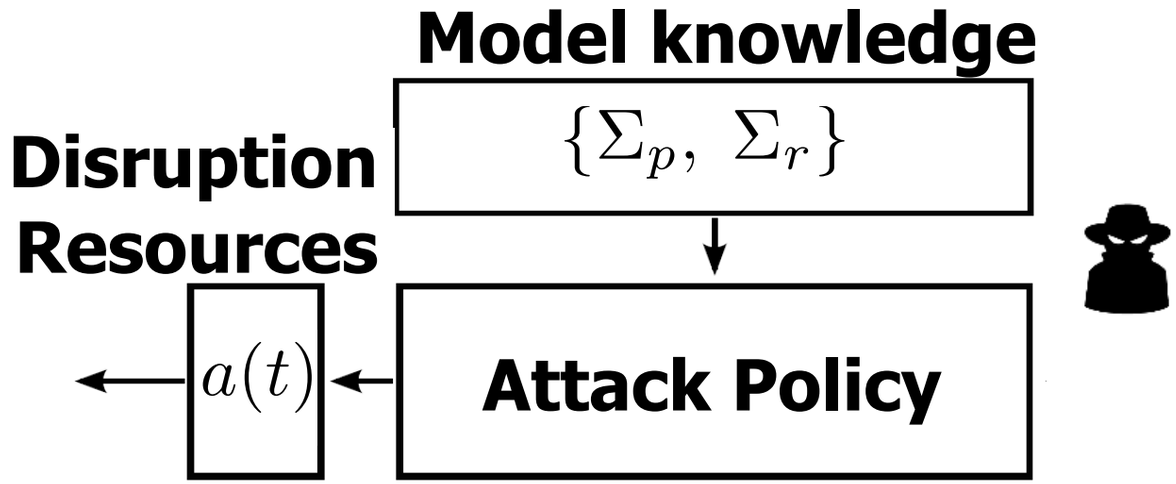
$$\Sigma_p = (A, B, C_p, D_p),\ G_p(s) = C_p(sI - A)^{-1}B + D_p$$
$$\Sigma_r = (A, B, C_r, D_r),\ G_r(s) = C_r(sI - A)^{-1}B + D_r$$

- **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



Adversary Model



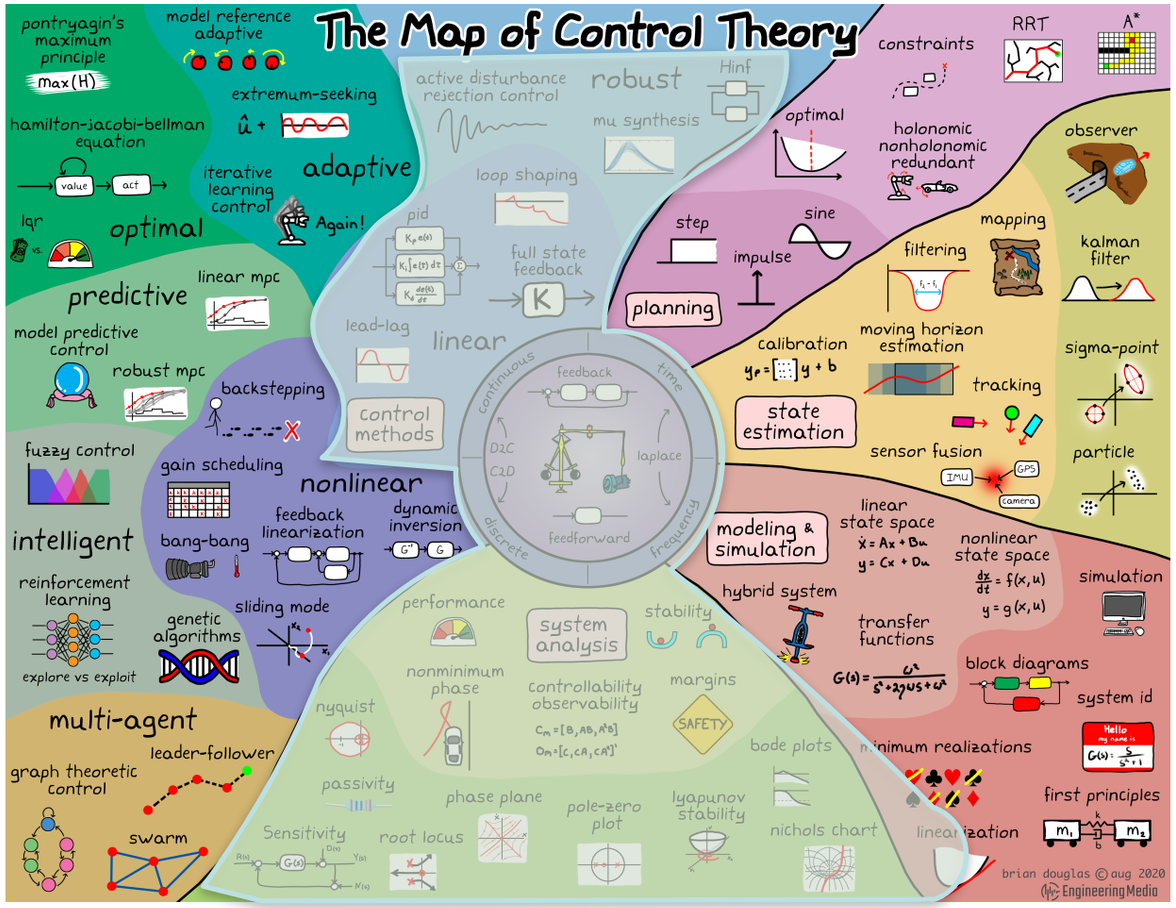


$$\Sigma_p = (A, B, C_p, D_p),\ G_p(s) = C_p(sI - A)^{-1}B + D_p$$
$$\Sigma_r = (A, B, C_r, D_r),\ G_r(s) = C_r(sI - A)^{-1}B + D_r$$

- **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?

Control System Design

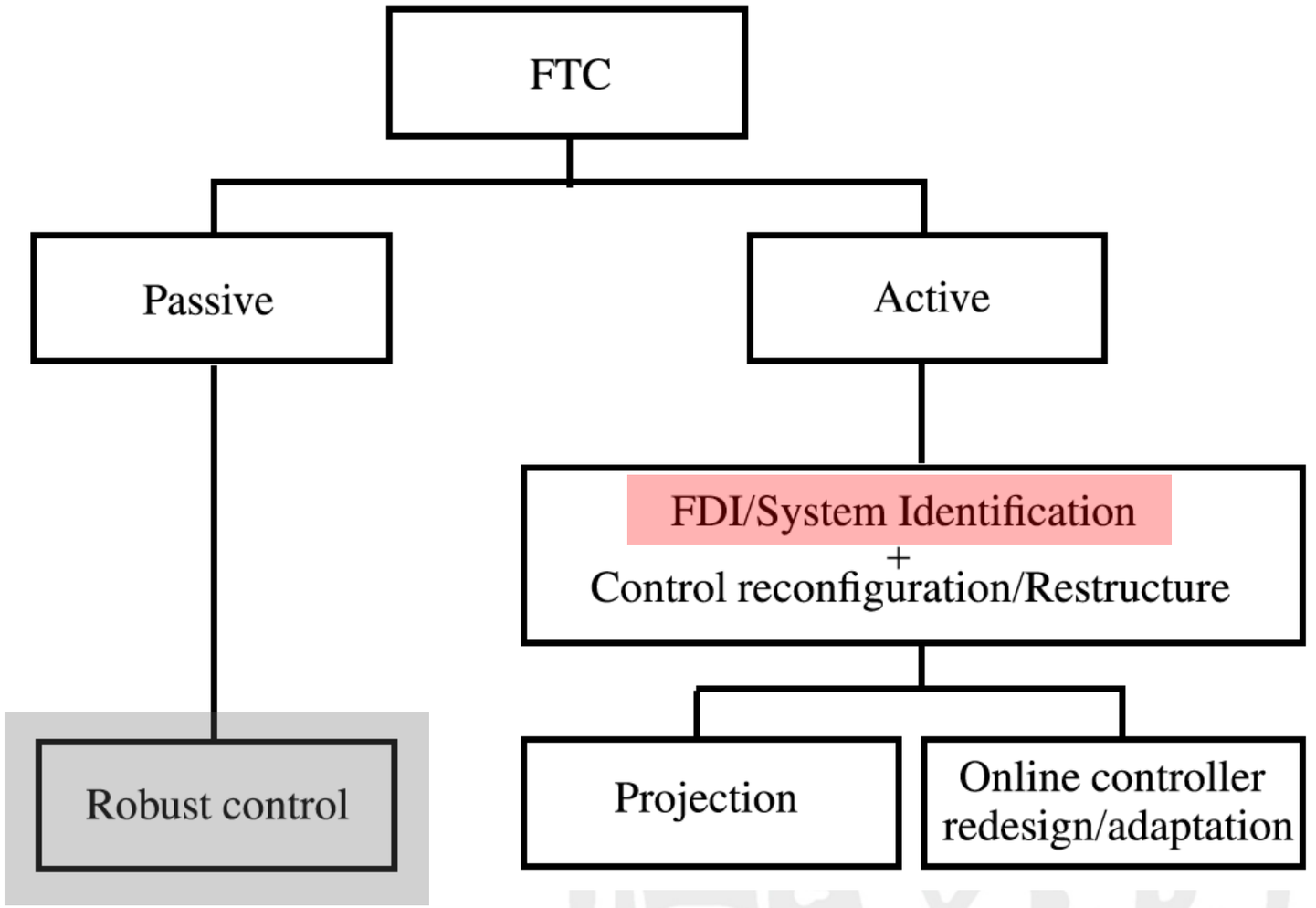


Main steps in control system design:

- Scenario characterization
  - Models, Scenarios, Cost functions
- System Analysis
  - Sensitivity metrics (w.r.t. input) [not binary in general]
- System Design (Control Methods)
  - Find control and monitoring algorithms that <u>minimise the sensitivity metric</u>
  - Re-design the structure of the system to <u>minimise the sensitivity metric</u>

Fault-Tolerant Control

- While faults are not *detected*:
  - *Robust controller*
- If a fault is *detected*:
  - adapt or redesign
- Key fields of interest for "weakly detectable" attacks:
  - Fault Detection
  - Robust Control



FTC design objectives

FTC design objectives

- **Robust controller design:** find a controller that
  - Minimizes the “worst-case” (largest) **impact** of unit-energy faults
  - i.e.: optimal H_∞ control

FTC design objectives

- **Robust controller design:** find a controller that
  - Minimizes the “worst-case” (largest) **impact** of unit-energy faults
  - i.e.: optimal H_∞ control

- **Fault detection filter design:** find an observer/filter that
  - Maximizes the “worst-case” (smallest) **detectability** of unit-energy faults
  - i.e.: optimal $H\_$ detection filter design

FTC design objectives

- **Robust controller design:** find a controller that
  - Minimizes the “worst-case” (largest) **impact** of unit-energy faults
  - i.e.: optimal H_∞ control

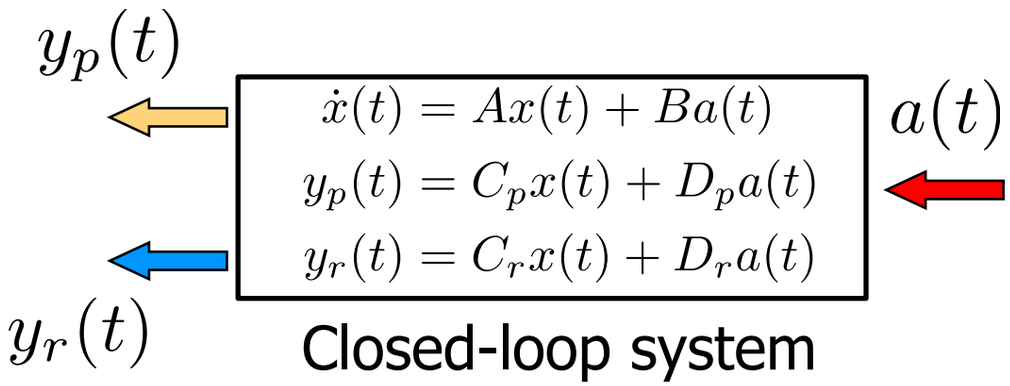
- **Fault detection filter design:** find an observer/filter that
  - Maximizes the “worst-case” (smallest) **detectability** of unit-energy faults
  - i.e.: optimal $H\_$ detection filter design

- Both are based on *sensitivity metrics*:
  - **Robustness**: largest **impact on performance** of unit-energy faults
  - **Detectability**: smallest **detectability** of unit-energy faults

Classical Sensitivity Metrics

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

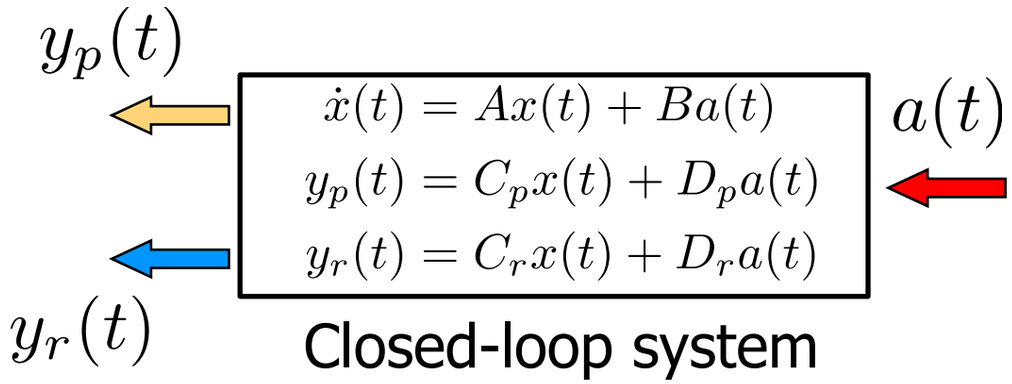
$$\|y\|_{\mathcal{L}_2}^2 \triangleq \int_{-\infty}^{+\infty} \|y(t)\|_2^2 \, dt$$



Classical Sensitivity Metrics

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

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



- **Robustness:**

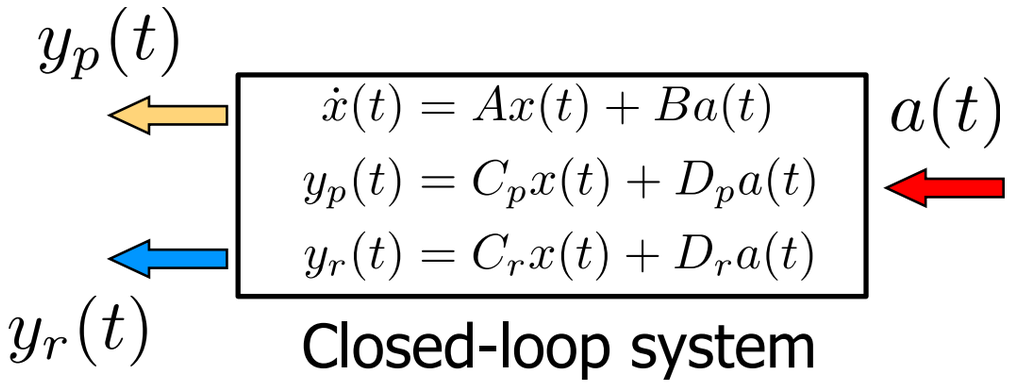
$$\gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0 \;\; (x(\infty) = 0)$$

Classical Sensitivity Metrics

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

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



- Robustness:

$$\gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$

$$x(0) = 0 \;\; (x(\infty) = 0)$$

- Frequency Domain:

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

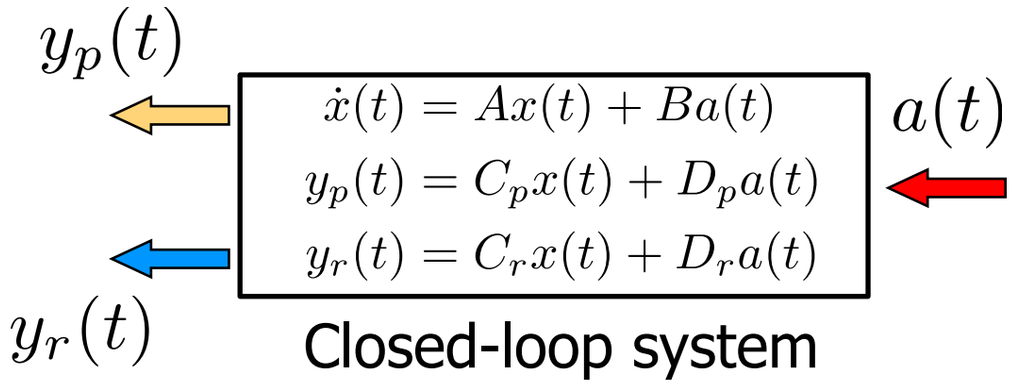
$$\bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^{n_a}} \|G_p(s)a\|_2$$

$$\text{s.t.} \quad \|a\|_2 = 1$$

Classical Sensitivity Metrics

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

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



- **Robustness:**

$$\gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$

$$x(0) = 0 \;\; (x(\infty) = 0)$$

- **Detectability:**

$$\gamma_{H_-} \triangleq \inf_{a \in \mathcal{L}_{2e}} \|y_r\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$

$$x(0) = 0 \quad x(\infty) = 0$$

- **Frequency Domain:**

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

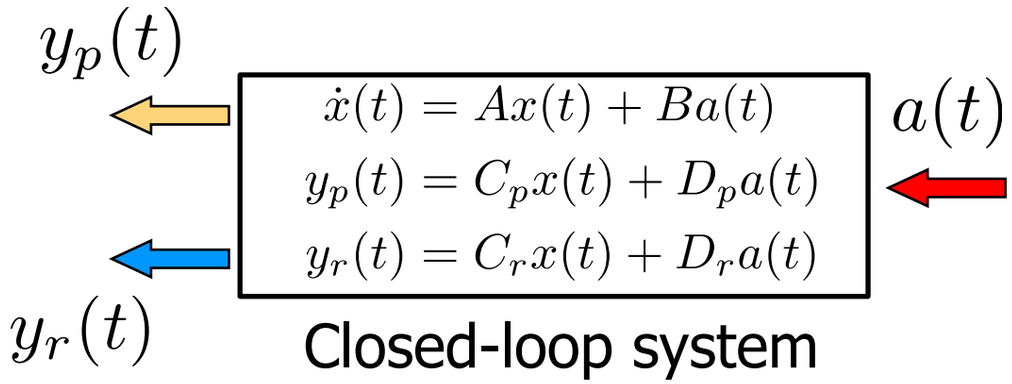
$$\bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^{n_a}} \|G_p(s)a\|_2$$

$$\text{s.t.} \quad \|a\|_2 = 1$$

Classical Sensitivity Metrics

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

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



- **Robustness:**

$$\gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$

$$x(0) = 0 \quad (x(\infty) = 0)$$

- **Frequency Domain:**

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

$$\bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^{n_a}} \quad \|G_p(s)a\|_2$$

$$\text{s.t.} \quad \|a\|_2 = 1$$

- **Detectability:**

$$\gamma_{H_-} \triangleq \inf_{a \in \mathcal{L}_{2e}} \quad \|y_r\|_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1$$

$$x(0) = 0 \quad x(\infty) = 0$$

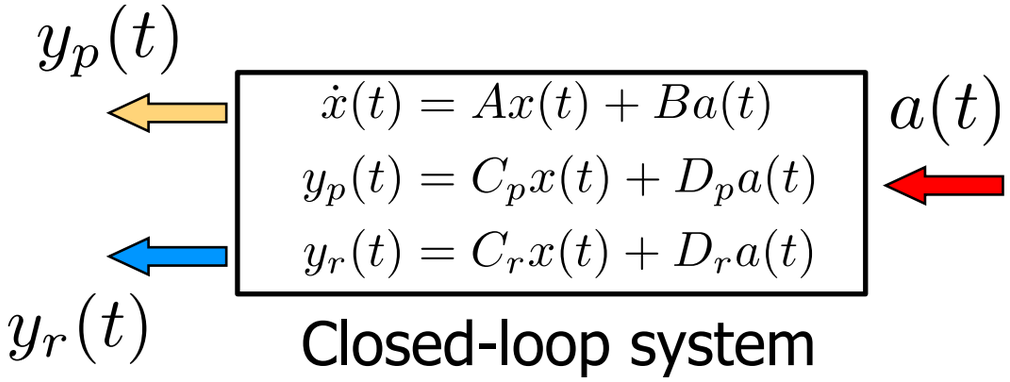
- **Frequency Domain:**

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

$$\underline{\sigma}_r(s) = \inf_{a \in \mathbb{C}^{n_a}} \quad \|G_r(s)a\|_2$$

$$\text{s.t.} \quad \|a\|_2 = 1$$

Classical Sensitivity Metrics

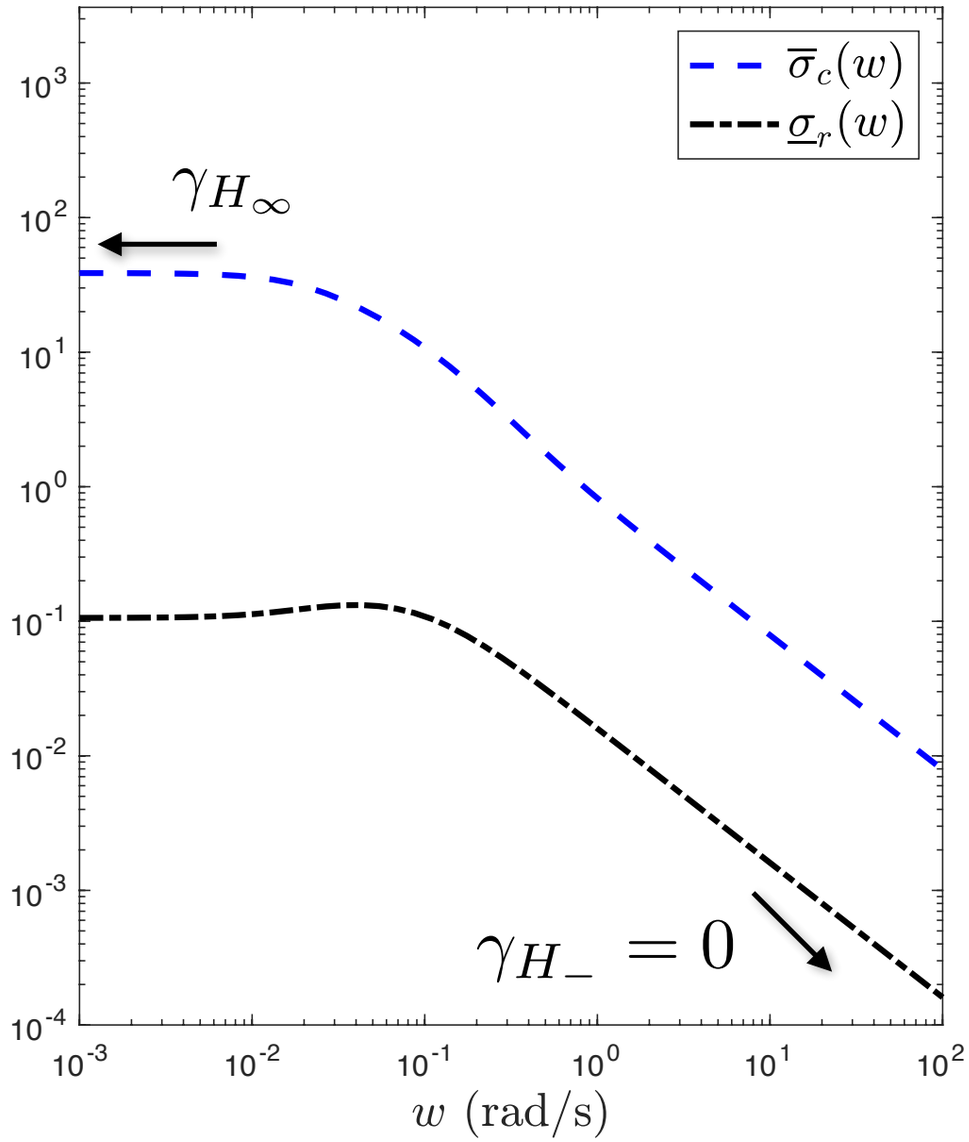


- Robustness (H_∞)

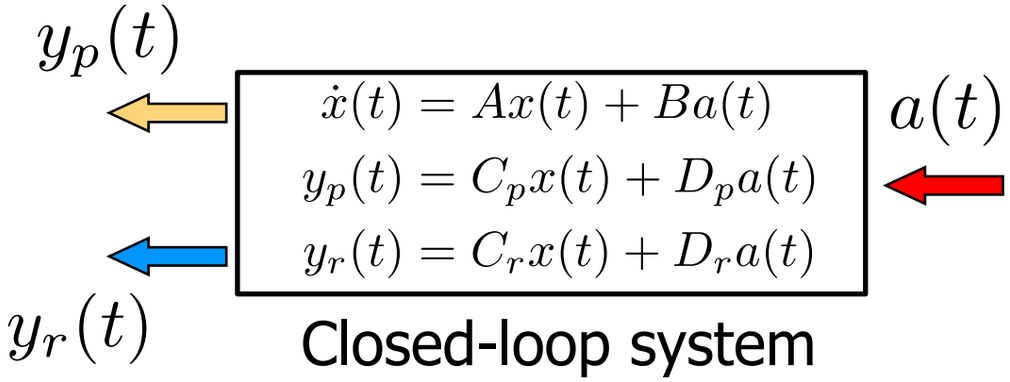
$$\gamma_{H_\infty} = \sup_{w \geq 0} \quad \bar{\sigma}_p(jw)$$

- Detectability ($H\_$)

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



Classical Sensitivity Metrics



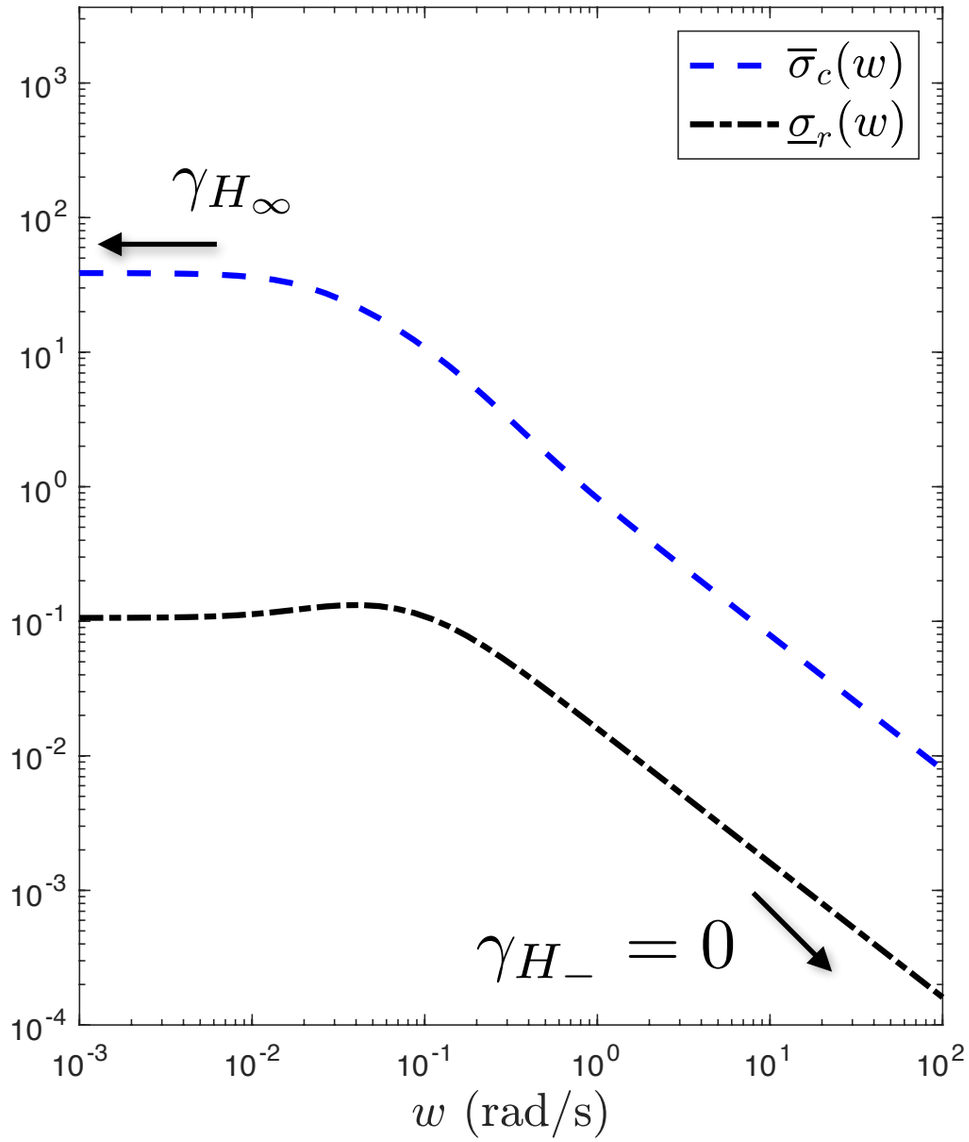
- Robustness (H_∞)

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

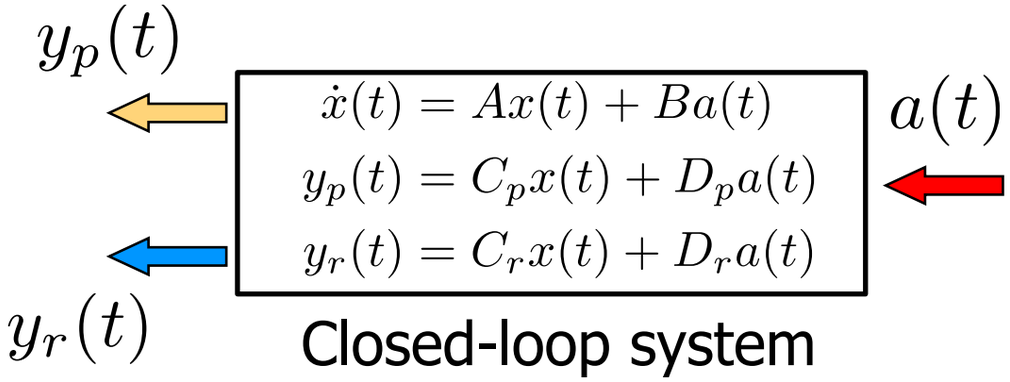
- Detectability (H_-)

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

- Least detectable fault has little impact…



Classical Sensitivity Metrics



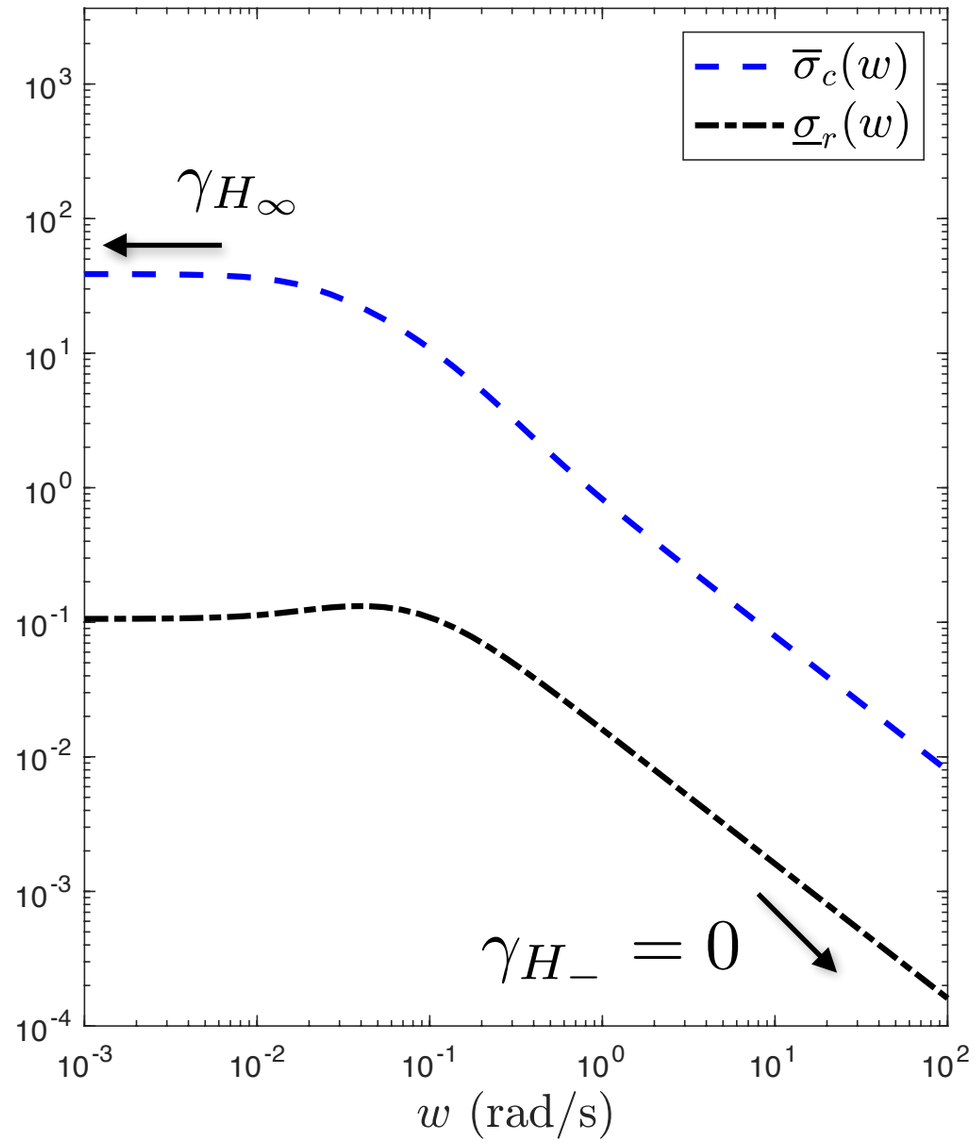
- Robustness (H_∞)

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

- Detectability (H_-)

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

- Least detectable fault has little impact…



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

Security metrics for control systems (a first heuristic)

- **Attack policy:** Maximise the impact on performance without raising alarms

Security metrics for control systems (a first heuristic)

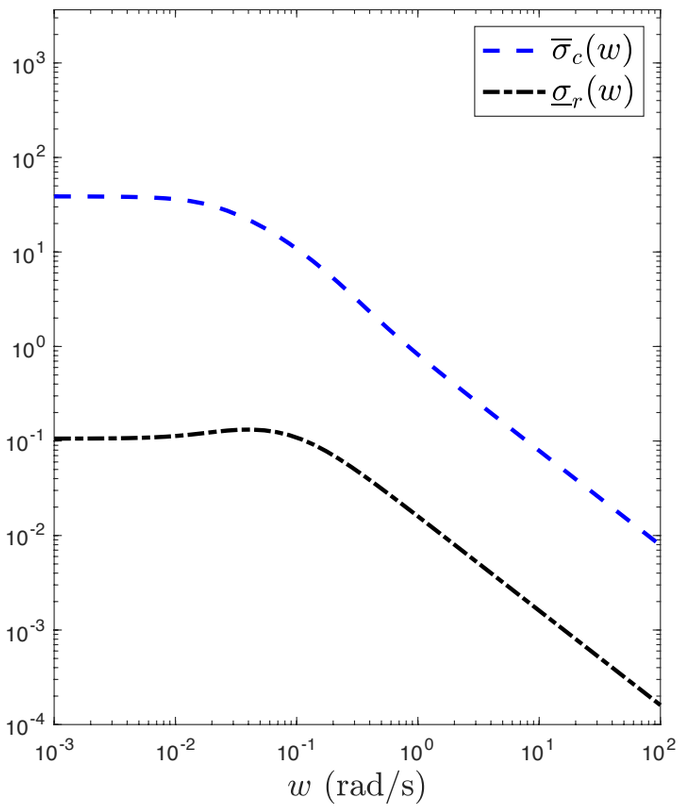
- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**

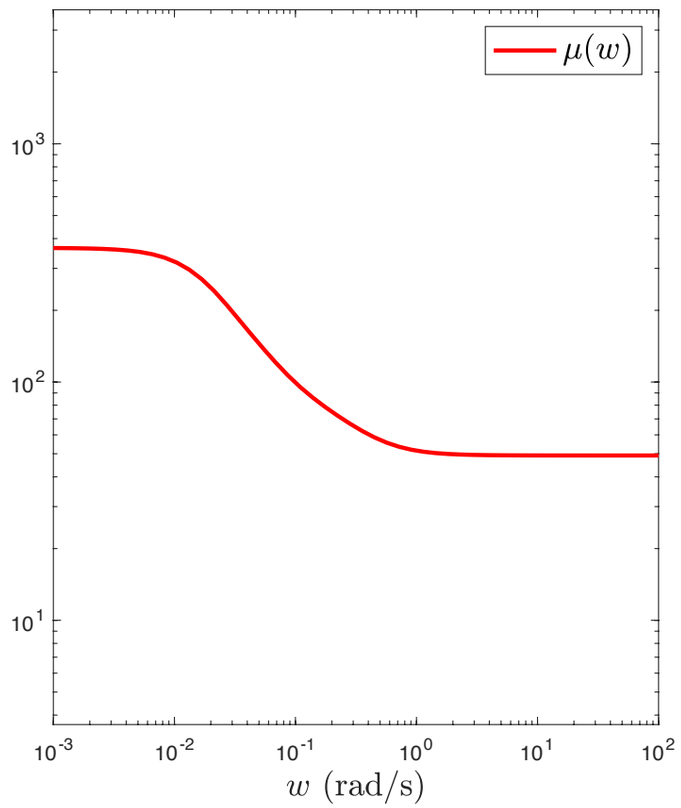
Security metrics for control systems (a first heuristic)

- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**
- "Desired" metric: combine H_∞ and $H\_$**:** $\mu(w) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

Security metrics for control systems (a first heuristic)

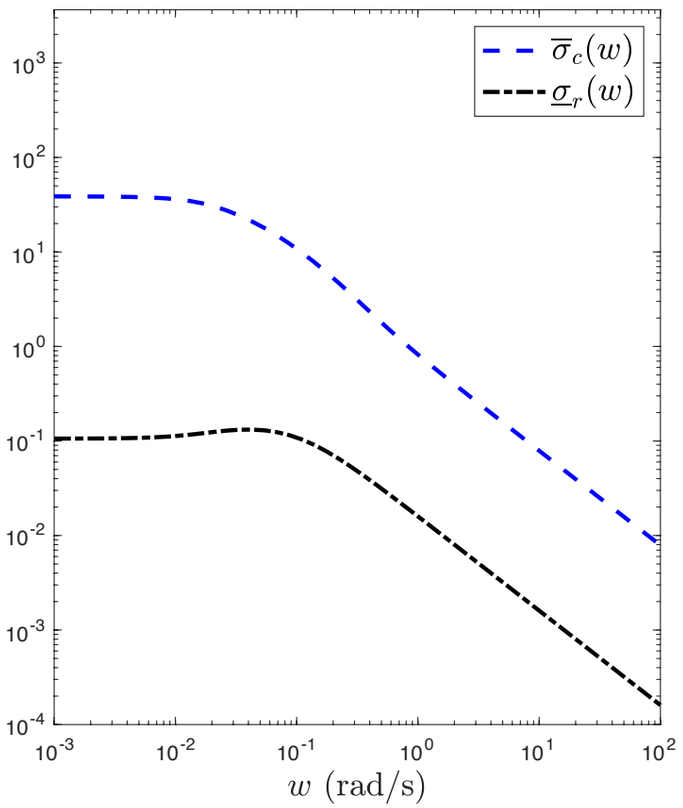
- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**
- "Desired" metric: combine H_∞ and $H\_$**:** $\mu(w) \triangleq \dfrac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

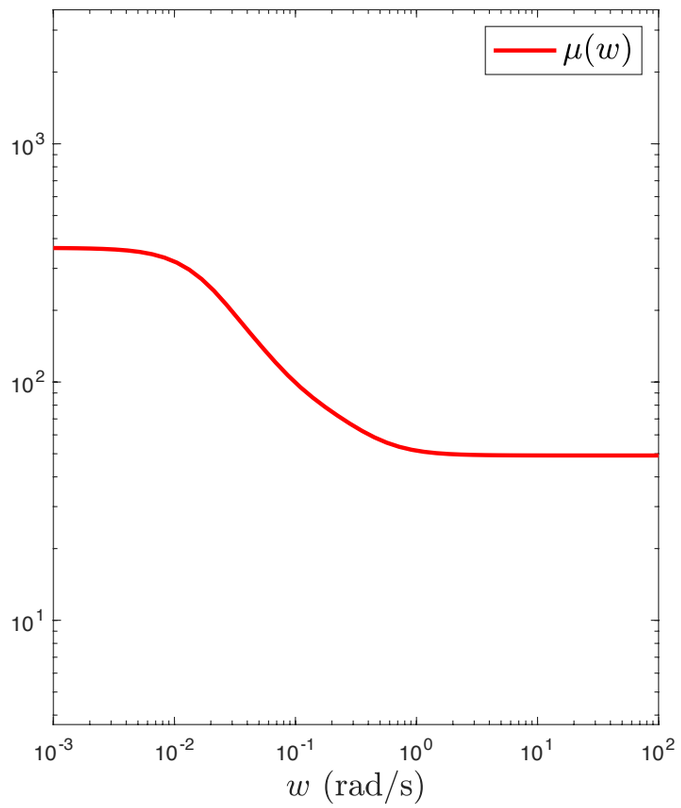




Security metrics for control systems (a first heuristic)

- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**
- "Desired" metric: combine H_∞ and $H\_$**:** $\mu(w) \triangleq \dfrac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

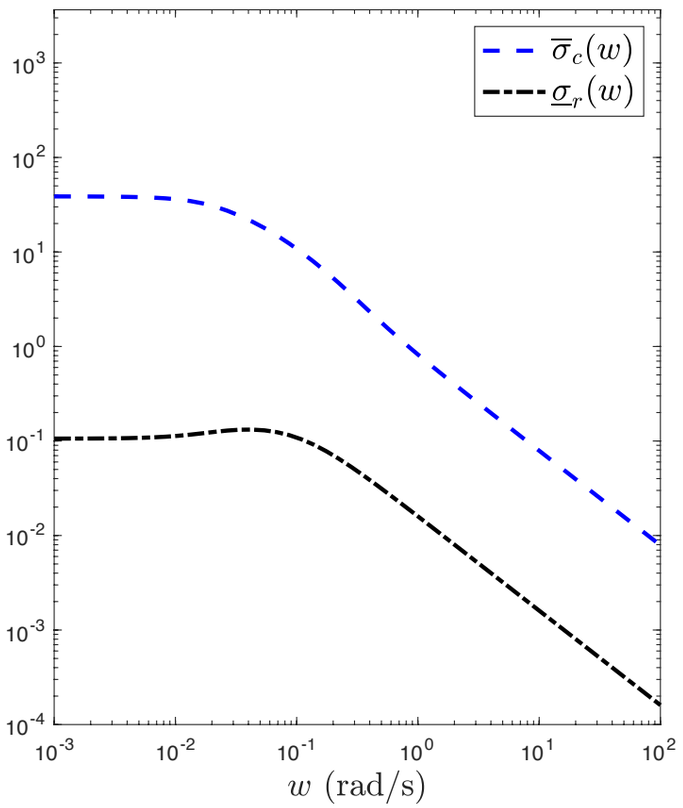


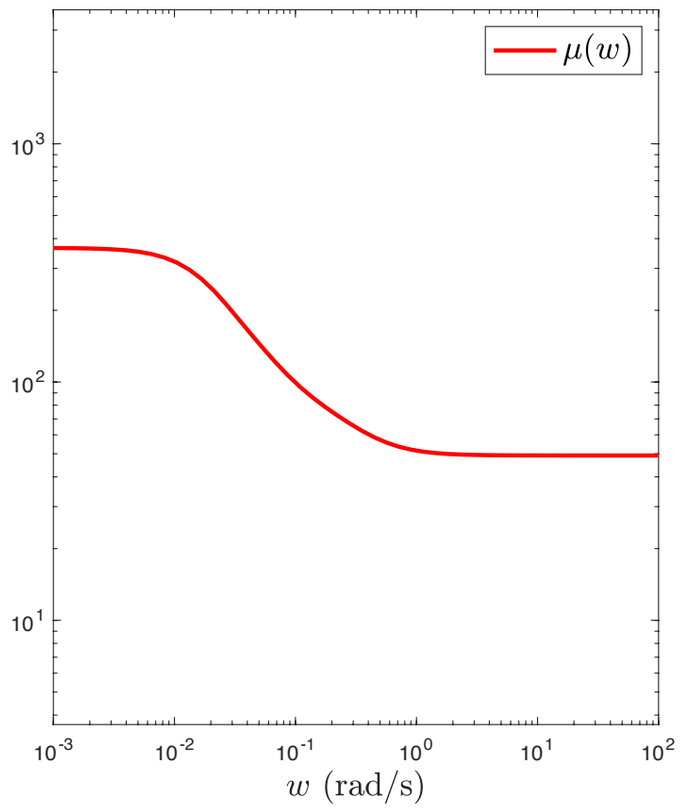


- But this is an heuristic…

Security metrics for control systems (a first heuristic)

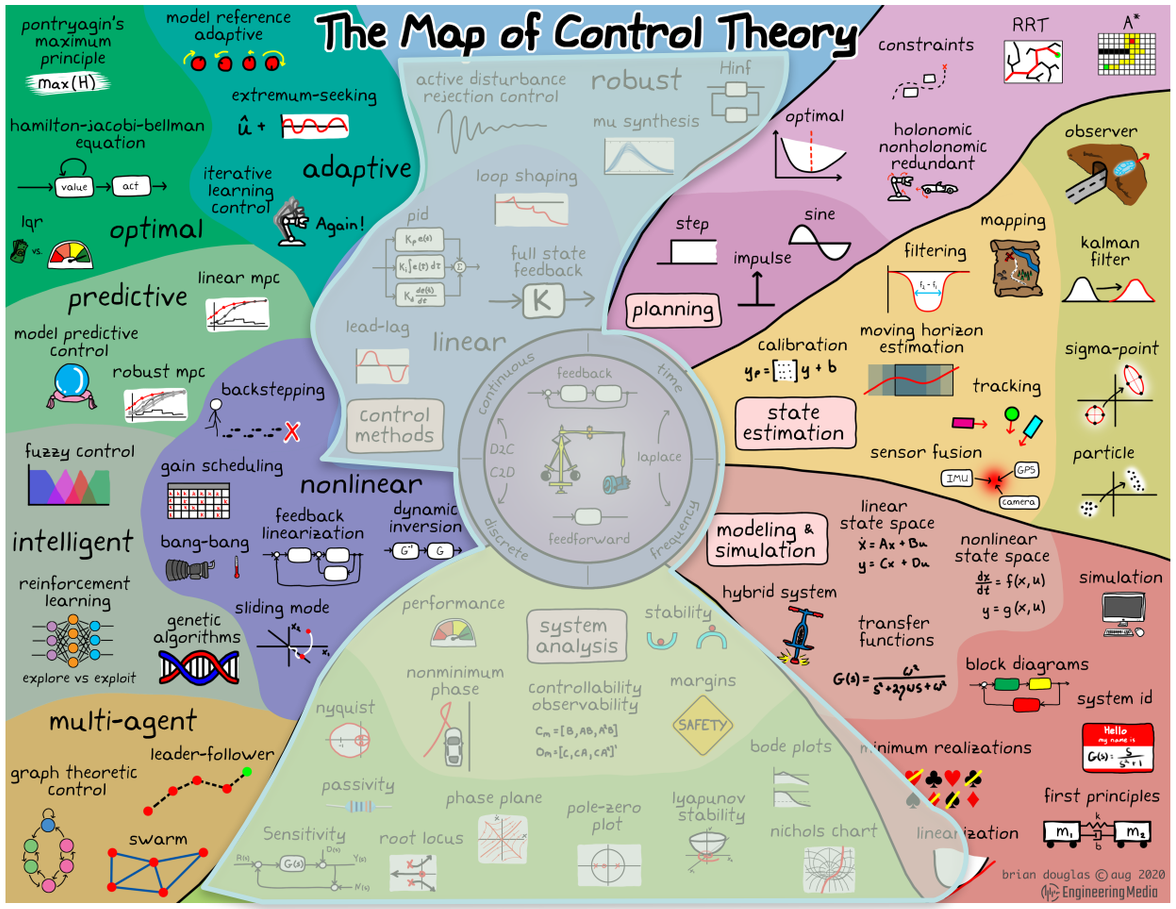
- **Attack policy:** Maximise the impact on performance without raising alarms
  - Requires a combination of **impact on performance** and **detection!**
- "Desired" metric: combine H_∞ and $H\_$**:** $\mu(w) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$





- But this is an heuristic…
- And how do we formalise such a metric?
  - … coming next!

Control System Design



Main steps in control system design:

- Scenario characterization
  - Models, Scenarios, Cost functions
- System Analysis
  - Sensitivity metrics (w.r.t. input)
  - **<u>Need to capture security aspects:</u>** maximise impact while undetected
- System Design (Control Methods)
  - Find control and monitoring algorithms that <u>minimise the sensitivity metric</u>
  - Re-design the structure of the system to <u>minimise the sensitivity metric</u>

Security metrics for control systems

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

Security metrics for control systems

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

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

[Teixeira et al., CDC 15]

Security metrics for control systems

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

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

[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"

Security metrics for control systems

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

- Non-convex infinite dimensional optimization problem
  - Contains “hidden convexity”

Security metrics for control systems

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

- Non-convex infinite dimensional optimization problem
  - Contains “hidden convexity”

- An equivalent formulation (dual problem):
  - Key step: Lossless S-procedure => Zero duality gap

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

Security metrics for control systems

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

- Non-convex infinite dimensional optimization problem
  - Contains “hidden convexity”

- An equivalent formulation (dual problem):
  - Key step: Lossless S-procedure => Zero duality gap

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

- Imposes a gain constraint on the outputs: $\gamma^{\star^2} \|y_r\|_{\mathcal{L}_2}^2 \geq \|y_p\|_{\mathcal{L}_2}^2$

Security metrics for control systems

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

- Non-convex infinite dimensional optimization problem
  - Contains “hidden convexity”

- An equivalent formulation (dual problem):
  - Key step: Lossless S-procedure => Zero duality gap

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

- Imposes a gain constraint on the outputs: $\gamma^{\star^2} \|y_r\|_{\mathcal{L}_2}^2 \geq \|y_p\|_{\mathcal{L}_2}^2$
- Finite $\gamma^\star$ implies a bound on the performance degradation by undetectable attacks

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

UPPSALA UNIVERSITET

Control System Analysis: Linear Matrix Inequalities

Control System Analysis: Linear Matrix Inequalities

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

- Infinite-dimensional constraint

Control System Analysis: Linear Matrix Inequalities

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

- Infinite-dimensional constraint

- An equivalent formulation:

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

[Teixeira, CDC 15]

Control System Analysis: Linear Matrix Inequalities

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

- Infinite-dimensional constraint

- An equivalent formulation:

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- This amounts to a convex SDP with Linear Matrix Inequalities (LMIs)

[Teixeira, CDC 15]

Control System Analysis: Linear Matrix Inequalities

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

- Infinite-dimensional constraint

- An equivalent formulation:

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- This amounts to a convex SDP with Linear Matrix Inequalities (LMIs)
- Can be efficiently solved by SDP solvers (e.g., through CVX)

[Teixeira, CDC 15]

Control System Analysis: Linear Matrix Inequalities

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

- Infinite-dimensional constraint

- An equivalent formulation:

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- This amounts to a convex SDP with Linear Matrix Inequalities (LMIs)
- Can be efficiently solved by SDP solvers (e.g., through CVX)
- Key technique: Dissipative Systems Theory

[Teixeira, CDC 15]

Dissipative Systems Theory

Consider the LTI system Σ with input a and outputs y_p and y_r. The following statements are equivalent:

1. the system Σ is dissipative w.r.t. $s(a,x) = \beta\|y_r(t)\|_2^2 - \|y_p(t)\|_2^2$;
2. for all trajectories of the system such that $T > 0$ and $x(0) = 0$, we have $\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$;
3. there exists a positive semi-definite matrix $P \succeq 0$ such that the following linear matrix inequality (LMI) holds:

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

J.C. Willems, “Dissipative dynamical systems Part II: Linear systems with quadratic supply rates”, Archive for Rational Mechanics and Analysis, 45 (5) (1972), pp.352-393

H.L. Trentelman, J.C. Willems, “The Dissipation Inequality and the Algebraic Riccati Equation”. In: Bittanti S., Laub A.J., Willems J.C. (eds) The Riccati Equation. Communications and Control Engineering Series. Springer, Berlin, Heidelberg (1991)

Dissipative Systems Theory

Consider the LTI system Σ with input a and outputs y_p and y_r. The following statements are equivalent:

1. the system Σ is dissipative w.r.t. $s(a,x) = \beta\|y_r(t)\|_2^2 - \|y_p(t)\|_2^2$;
2. for all trajectories of the system such that $T > 0$ and $x(0) = 0$, we have $\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$;
3. there exists a positive semi-definite matrix $P \succeq 0$ such that the following linear matrix inequality (LMI) holds:

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

J.C. Willems, “Dissipative dynamical systems Part II: Linear systems with quadratic supply rates”, Archive for Rational Mechanics and Analysis, 45 (5) (1972), pp.352-393

H.L. Trentelman, J.C. Willems, “The Dissipation Inequality and the Algebraic Riccati Equation”. In: Bittanti S., Laub A.J., Willems J.C. (eds) The Riccati Equation. Communications and Control Engineering Series. Springer, Berlin, Heidelberg (1991)

- *Note of caution:* in general, there is no simple equivalent frequency domain inequality

H.L. Trentelman. When does the algebraic Riccati equation have a negative semi-definite solution?. In: Blondel, V., Sontag, E.D., Vidyasagar, M., Willems, J.C. (eds) Open Problems in Mathematical Systems and Control Theory. Communications and Control Engineering. Springer (1999)

On the Frequency Domain Inequality (extra)

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

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

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?
  - Not in general, but there is a necessary condition for the LMI to be feasible

Necessary condition: $\beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$ for all $s \in \mathbb{C}$ with $s \notin \sigma(A)$, $\mathrm{Re}(s) \geq 0$.

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?
  - Not in general, but there is a necessary condition for the LMI to be feasible

Necessary condition: $\beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$ for all $s \in \mathbb{C}$ with $s \notin \sigma(A)$, $\mathrm{Re}(s) \geq 0$.

- Not sufficient, see a counter example in J. C. Willems, "On the existence of a nonpositive solution to the Riccati equation", IEEE Trans. Automat. Contr., vol. AC-19, pp. 592-593, Oct. 1974.

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?
  - Not in general, but there is a necessary condition for the LMI to be feasible

Necessary condition: $\beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$ for all $s \in \mathbb{C}$ with $s \notin \sigma(A)$, $\mathrm{Re}(s) \geq 0$.

- Not sufficient, see a counter example in J. C. Willems, "On the existence of a nonpositive solution to the Riccati equation", IEEE Trans. Automat. Contr., vol. AC-19, pp. 592-593, Oct. 1974.
- But special cases do have an equivalent FDI:
  - see H_∞ norm (and the “Periodic Attacks” in later slides).

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?
  - Not in general, but there is a necessary condition for the LMI to be feasible

Necessary condition: $\beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$ for all $s \in \mathbb{C}$ with $s \notin \sigma(A)$, $\mathrm{Re}(s) \geq 0$.

- Not sufficient, see a counter example in J. C. Willems, "On the existence of a nonpositive solution to the Riccati equation", IEEE Trans. Automat. Contr., vol. AC-19, pp. 592-593, Oct. 1974.
- But special cases do have an equivalent FDI:
  - see H_∞ norm (and the “Periodic Attacks” in later slides).
- Necessary condition relates to generalised singular values: $\mu(s) \triangleq \min_{\beta} \quad \sqrt{\beta}$

$$\text{s.t.} \quad \beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$$

On the Frequency Domain Inequality (extra)

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

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

- Is there a frequency domain inequality (FDI) equivalent to the LMI?
  - Not in general, but there is a necessary condition for the LMI to be feasible

Necessary condition: $\beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$ for all $s \in \mathbb{C}$ with $s \notin \sigma(A)$, $\mathrm{Re}(s) \geq 0$.

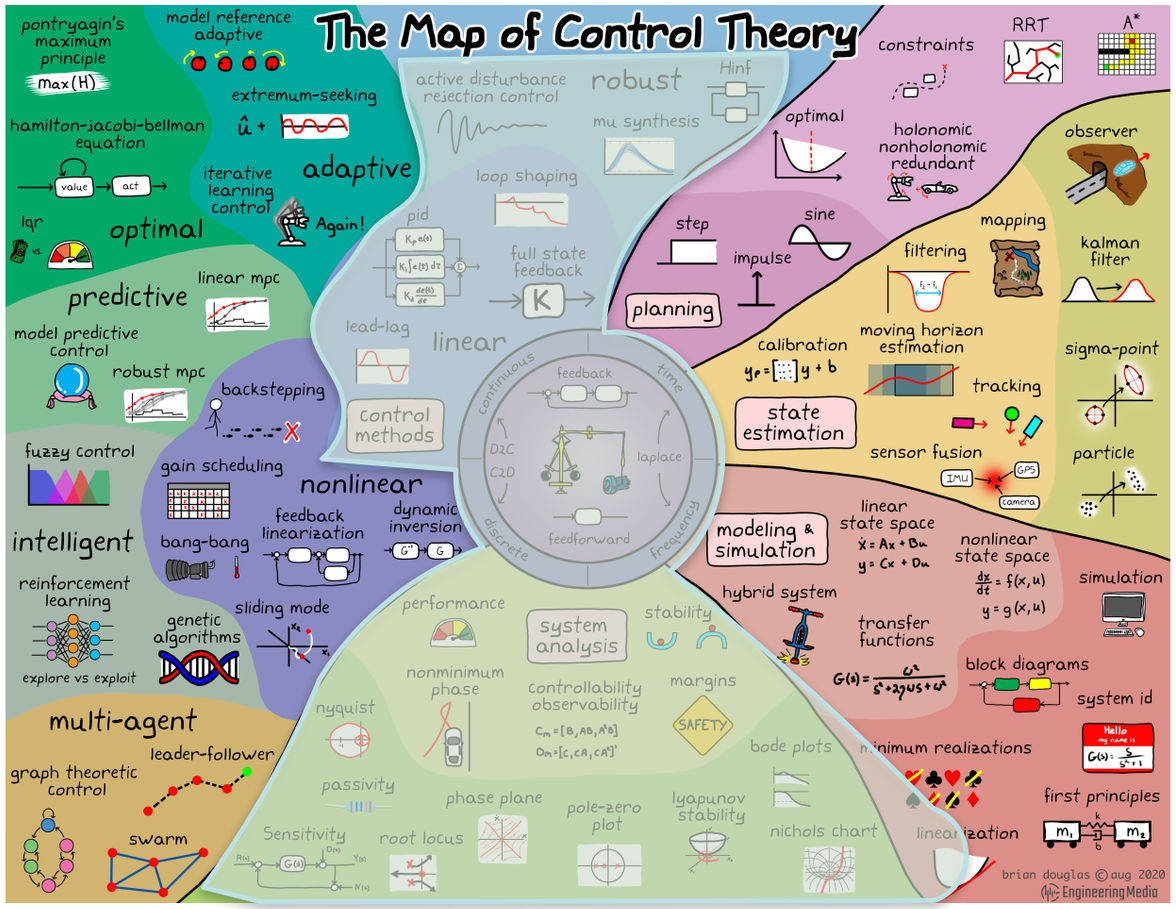
- Not sufficient, see a counter example in J. C. Willems, "On the existence of a nonpositive solution to the Riccati equation", IEEE Trans. Automat. Contr., vol. AC-19, pp. 592-593, Oct. 1974.
- But special cases do have an equivalent FDI:
  - see H_∞ norm (and the “Periodic Attacks” in later slides).
- Necessary condition relates to generalised singular values: $\mu(s) \triangleq \min_\beta \quad \sqrt{\beta}$

$$\text{s.t.} \quad \beta G_r(\bar{s})^\top G_r(s) - G_p(\bar{s})^\top G_p(s) \succeq 0$$

- For a single attack channel (input), we get: $\mu(s) = \dfrac{\overline{\sigma}_p(s)}{\underline{\sigma}_r(s)}$

$$\begin{aligned} \mu(s) \triangleq \min_\beta \quad & \sqrt{\beta} \\ \text{s.t.} \quad & \beta g_r(\bar{s})^\top g_r(s) - g_p(\bar{s})^\top g_p(s) \succeq 0 \\ = \min_\beta \quad & \sqrt{\beta} \\ \text{s.t.} \quad & \beta \underline{\sigma}_r(s)^2 - \overline{\sigma}_p(s)^2 \geq 0 \end{aligned}$$

Control System Design



Main steps in control system design:

- Scenario characterization
  - Models, Scenarios, Cost functions
- System Analysis
  - Sensitivity metrics (w.r.t. input)
  - **<u>Need to capture security aspects:</u>** maximise impact while undetected
- System Design (Control Methods)
  - Find control and monitoring algorithms that <u>minimise the sensitivity metric</u>
  - Re-design the structure of the system to <u>minimise the sensitivity metric</u>

Controller and Detector Design: Bilinear Matrix Inequalities

- **Design problem for** a Controller (L) and a Detector (K):
  - K and L change the matrices of the closed-loop system

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

[Teixeira, Springer 2021]

Controller and Detector Design: Bilinear Matrix Inequalities

- **Design problem for** a Controller (L) and a Detector (K):
  - K and L change the matrices of the closed-loop system

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

- **Designing** a Controller (L) & Detector (K) is possible under specific forms
  - … but leads to Bilinear Matrix Inequalities

[Teixeira, Springer 2021]

$$\min_{P \succeq 0, \beta > 0, K, L} \quad \beta$$

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

Controller and Detector Design: Bilinear Matrix Inequalities

- **Design problem for** a Controller (L) and a Detector (K):
  - K and L change the matrices of the closed-loop system

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

- **Designing** a Controller (L) & Detector (K) is possible under specific forms
  - … but leads to Bilinear Matrix Inequalities

[Teixeira, Springer 2021]

$$\min_{P \succeq 0, \beta > 0, K, L} \quad \beta$$

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

- Next we use an heuristic: alternating minimisation
  - Fix K,L, β, and solve for P
  - Fix P, solve for K,L, and β.

Example 1: Continuous-time

[Teixeira, Springer 2021]

- Classical vs Re-designed controller and detector

Nominal Design

After re-design

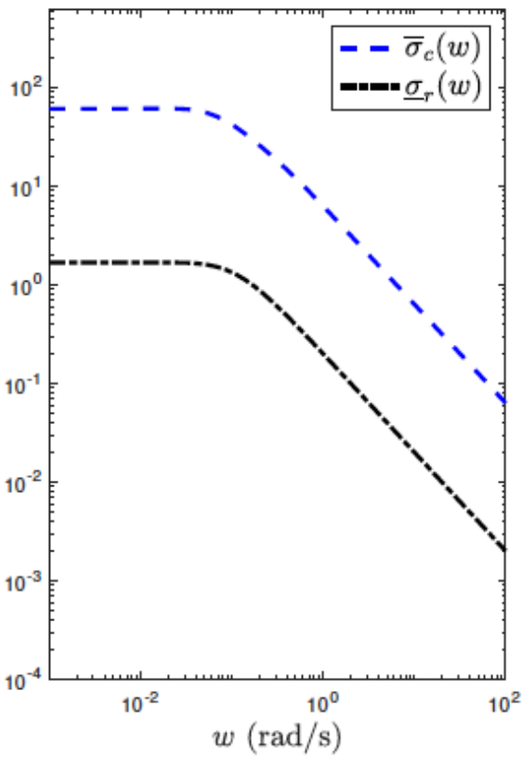
$$\mu(w) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

Example 1: Continuous-time

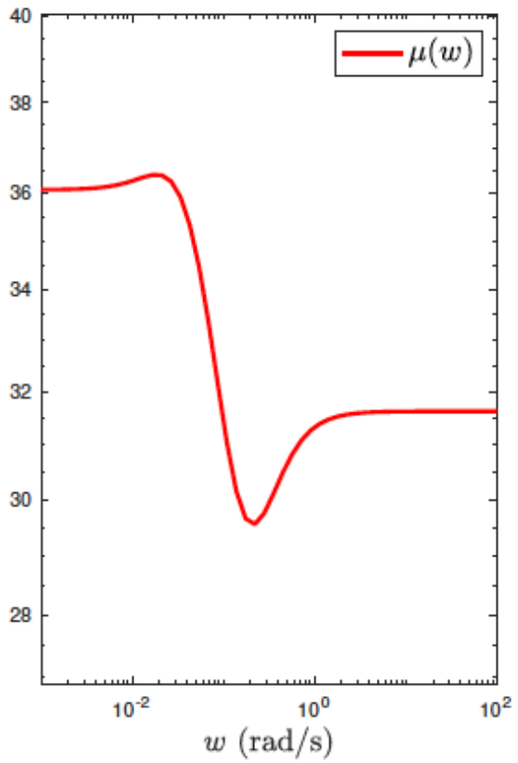
[Teixeira, Springer 2021]

- Classical vs Re-designed controller and detector

Nominal Design



After re-design



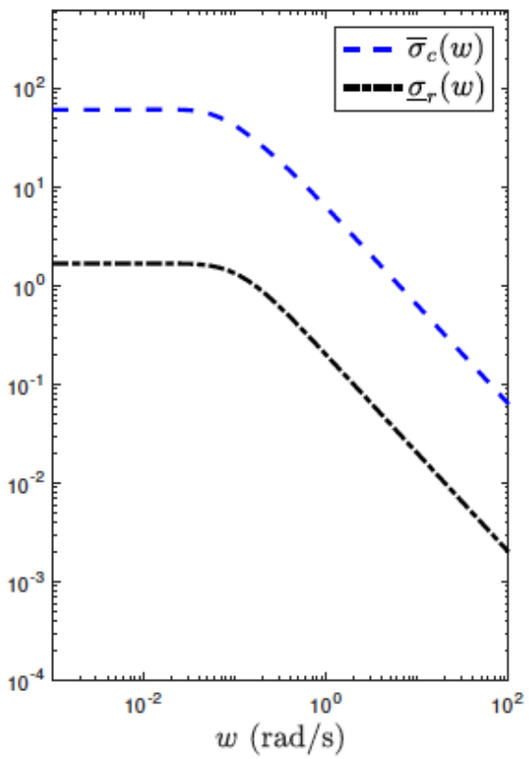
$$\mu(w) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

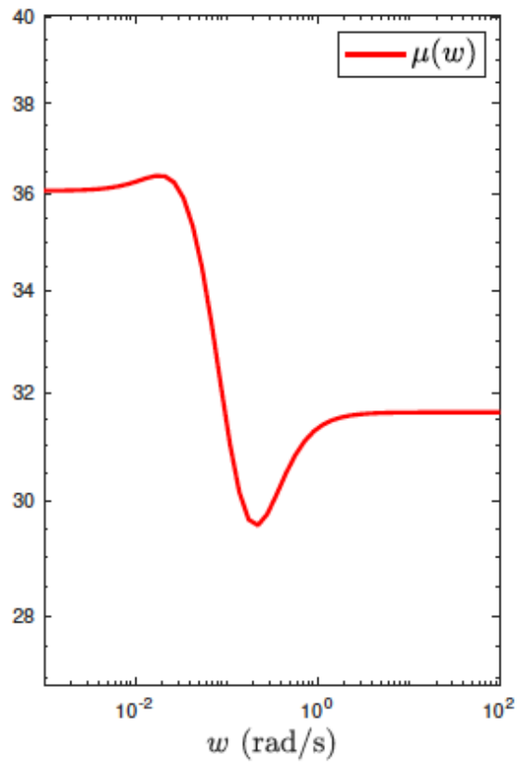
Example 1: Continuous-time

[Teixeira, Springer 2021]

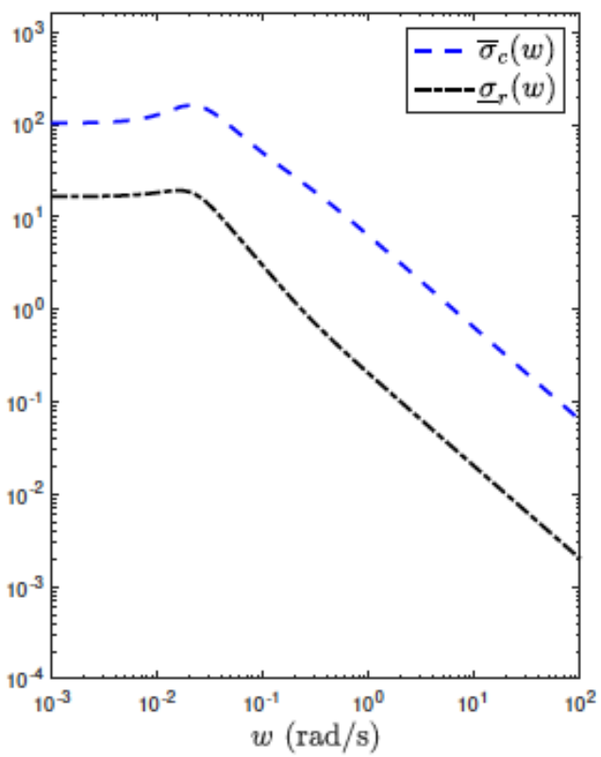
- Classical vs Re-designed controller and detector

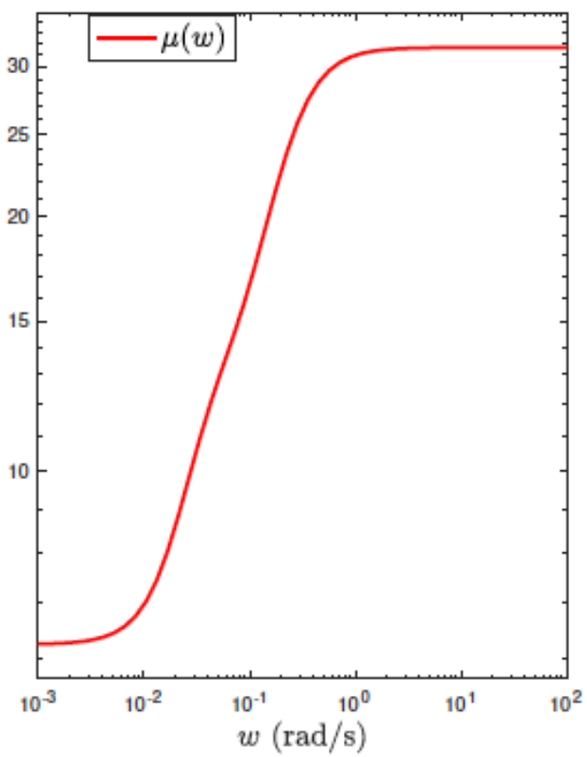
Nominal Design





After re-design





$$\mu(w) \triangleq \frac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$$

Example 2: Discrete-time

[Anand and Teixeira, IFAC 2020]

- Classical vs Re-designed controller and detector

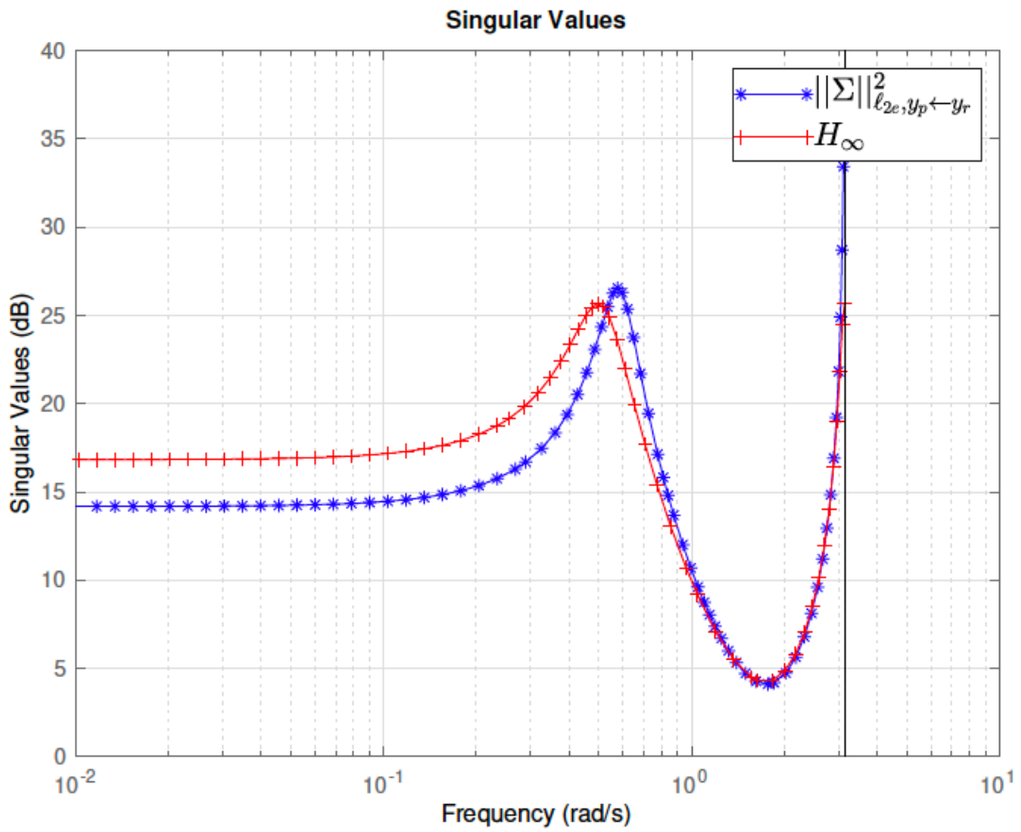


Fig. 1. Singular values - Performance output ($\bar{\sigma}(\Sigma_p)$)

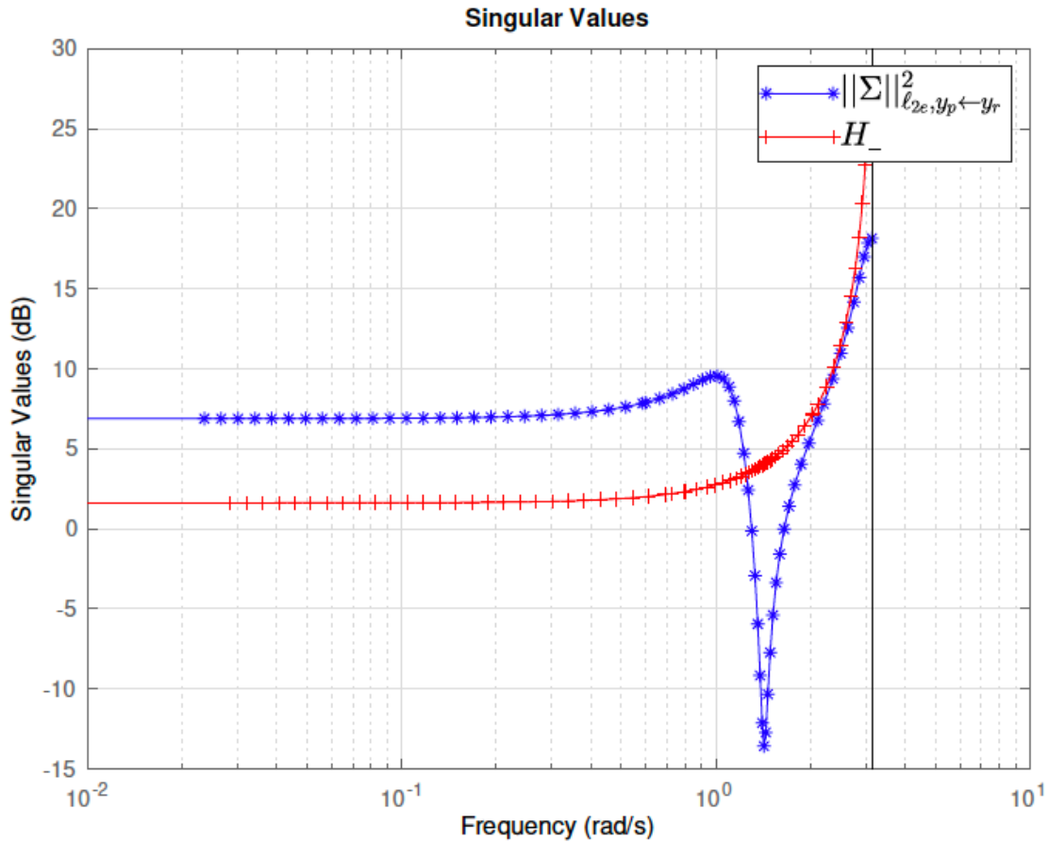


Fig. 2. Singular values - Detection output ($\underline{\sigma}(\Sigma_r)$)

UPPSALA UNIVERSITET

Security Metrics for Control Systems: from Analysis to Design

Security Metrics for Control Systems: from Analysis to Design

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

- Non-convex infinite dimensional optimization problem

Security Metrics for Control Systems: from Analysis to Design

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

- Non-convex infinite dimensional optimization problem

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- Contains “hidden convexity”

Security Metrics for Control Systems: from Analysis to Design

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

- Non-convex infinite dimensional optimization problem

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- Contains “hidden convexity”

$$\min_{K,L} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_p\|_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0$$

- Min-max approach to algorithm design

Security Metrics for Control Systems: from Analysis to Design

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

- Non-convex infinite dimensional optimization problem

$$\gamma^{\star^2} = \min_{\beta \geq 0,\, P \succeq 0} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A^\top P + PA & PB \\ B^\top P & 0 \end{bmatrix} - \beta \begin{bmatrix} C_r^\top C_r & C_r^\top D_r \\ D_r^\top C_r & D_r^\top D_r \end{bmatrix} + \begin{bmatrix} C_p^\top C_p & C_p^\top D_p \\ D_p^\top C_p & D_p^\top D_p \end{bmatrix} \preceq 0$$

- Contains “hidden convexity”

$$\min_{K,L} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_p\|_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0$$

- Min-max approach to algorithm design

$$\min_{P \succeq 0, \beta > 0, K, L} \quad \beta$$
$$\text{s.t.} \quad \begin{bmatrix} A(K,L)^\top P + PA(K,L) & PB(K,L) & C_p(K,L)^\top \\ B(K,L)^\top P & 0 & D_p(K,L)^\top \\ C_p(K,L) & D_p(K,L) & \beta I \end{bmatrix} - \beta \begin{bmatrix} C_r \top \\ D_r^\top \\ 0 \end{bmatrix} \begin{bmatrix} C_r & D_r & 0 \end{bmatrix} \preceq 0,$$

- Can be cast as an SDP problem with BMIs

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks

But when is the gain constraint infeasible?

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

- **Undetectable attacks:**

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

- **Undetectable attacks:**
  - Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

- **Undetectable attacks:**
  - Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  - Attack policy: $a_k = \nu^k g$
    - $|\nu| < 1$: vanishing attack
    - $|\nu| > 1$: diverging attack

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

- **Undetectable attacks:**
  - Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

  - Attack policy: $a_k = \nu^k g$
    - $|\nu| < 1$: vanishing attack
    - $|\nu| > 1$: diverging attack
  - Detection output can be made **arbitrarily small**, while the **performance output diverges.**

Structural Limitations

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

$$\int_0^T \|y_p(t)\|_2^2 \leq \beta \int_0^T \|y_r(t)\|_2^2$$

Finite $\gamma^\star$ implies bound on performance degradation by undetectable attacks
But when is the gain constraint infeasible?

Answer: When there are unstable invariant zeros from $a(t)$ to $y_r(t)$ that are not zeros from $a(t)$ to $y_p(t)$.

- **Undetectable attacks:**
  - Zero dynamics:

$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

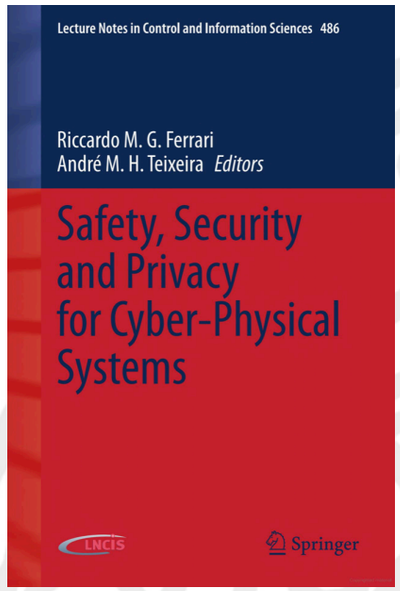
  - Attack policy: $a_k = \nu^k g$
    - $|\nu| < 1$: vanishing attack
    - $|\nu| > 1$: diverging attack
  - Detection output can be made **arbitrarily small**, while the **performance output diverges.**
- **Relative degree:** "zeros at infinite frequencies"
  - relates to rate of decay of singular values at infinite frequency
  - Infeasible gain constraint if $\underline{\sigma}_r(jw)$ decays faster than $\overline{\sigma}_p(jw)$

Outline

- Cybersecurity in Control Systems
  - Security Game
  - Risk Management
    - Mapping to Control System Design
    - Example: undetectable attacks

- Security Metrics for Control Systems
  - Classical metrics in control engineering
  - Novel security metric for analysis and design
    - Analysis and design problems
    - Structural limitations - invariant zeros

- Additional topics
  - Variations of the adversary models and their respective metrics
  - Incorporating uncertainty for robust open-loop attacks



Variations of adversary models

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

- Metrics for other scenarios can be formulated in a similar fashion, with additional constraints

* part of Tutorial 1

Variations of adversary models

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

- Periodic attacks*

$$\sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0,\ x(\infty) = 0$$

- Metrics for other scenarios can be formulated in a similar fashion, with additional constraints

* part of Tutorial 1

Variations of adversary models

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

- Periodic attacks*

$$\sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0,\ x(\infty) = 0$$

- FDI (on j axis) is equivalent to LMI (with symmetric P)

- Metrics for other scenarios can be formulated in a similar fashion, with additional constraints

* part of Tutorial 1

Variations of adversary models

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

- Periodic attacks*

$$\sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1$$
$$x(0) = 0,\ x(\infty) = 0$$

- FDI (on j axis) is equivalent to LMI (with symmetric P)
- $\gamma^\star = \sup_{\omega} \ \mu(w)$ (see slide 31 on FDI)
  - For single input case*: $\mu(w) \triangleq \dfrac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

- Metrics for other scenarios can be formulated in a similar fashion, with additional constraints

* part of Tutorial 1

Variations of adversary models

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

- Metrics for other scenarios can be formulated in a similar fashion, with additional constraints

- Periodic attacks*

$$\sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2} \quad \text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1, \; x(0) = 0, \; x(\infty) = 0$$

- FDI (on j axis) is equivalent to LMI (with symmetric P)
- $\gamma^\star = \sup_{\omega} \; \mu(w)$ (see slide 31 on FDI)
  - For single input case*: $\mu(w) \triangleq \dfrac{\bar{\sigma}_p(jw)}{\underline{\sigma}_r(jw)}$

- Truncated attacks

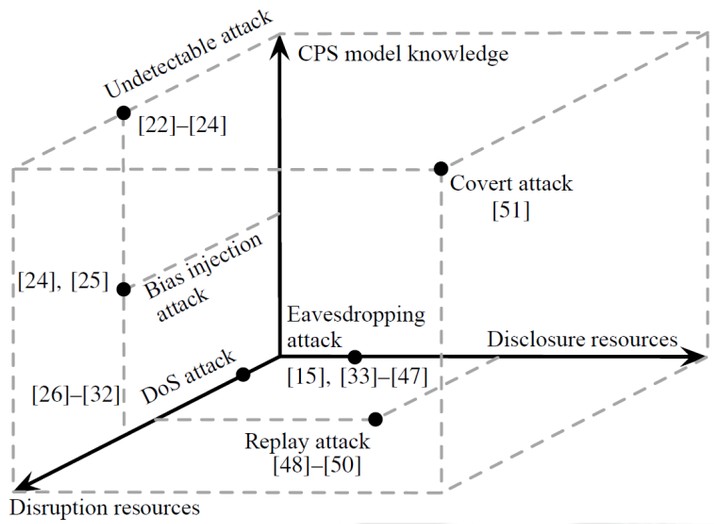
$$\sup_{a \in \mathcal{L}_{2e}} \|y_p\|_{\mathcal{L}_2} \quad \text{s.t.} \quad \|y_r\|_{\mathcal{L}_2} = 1, \; x(-\infty) = 0, \; x(\infty) = 0, \; a(t) = 0, \; \forall t \geq 0$$

[Teixeira, CDC 19]

* part of Tutorial 1

Structural limitation: a closer look

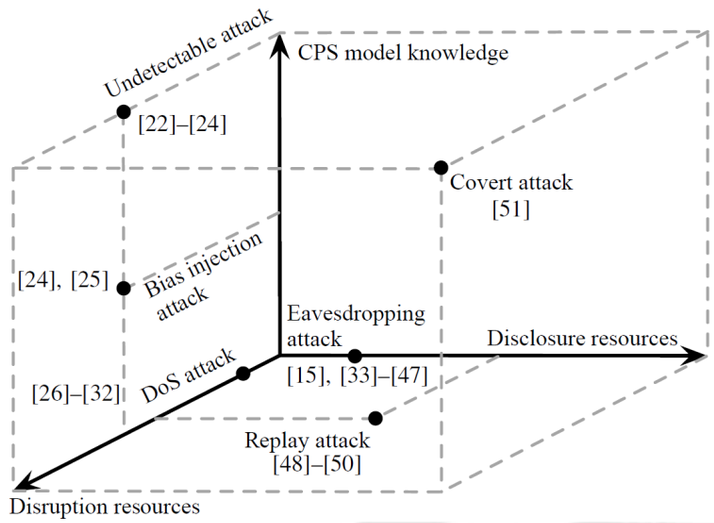
Structural limitation: a closer look



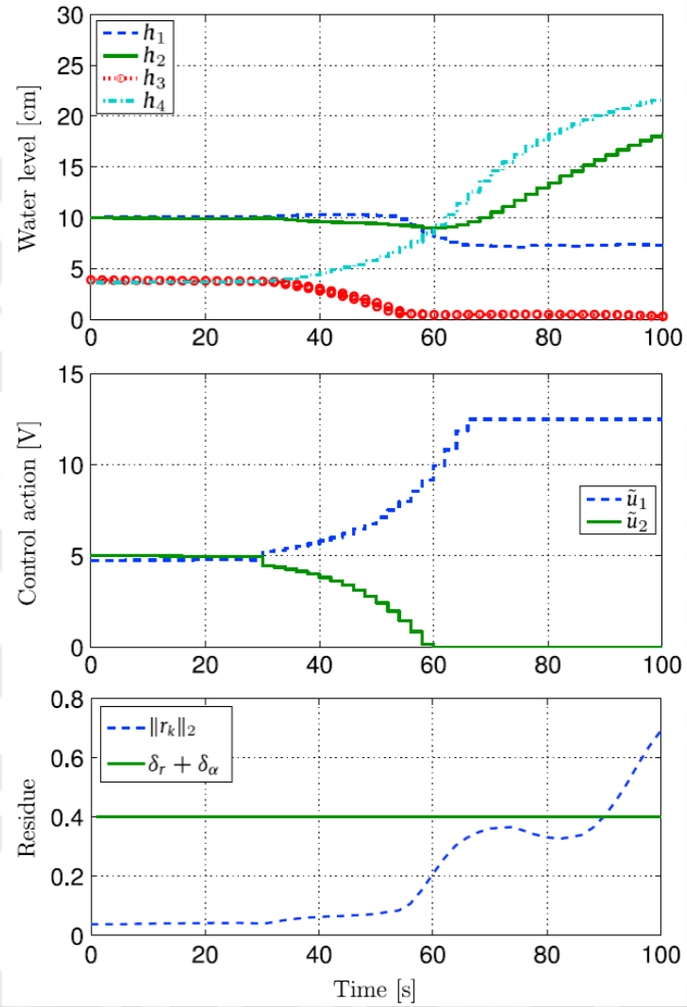
- Fundamental structural limitations still exist
  - Optimal attacks can give unbounded impact while undetected, regardless of the controller and detector
  - Related to undetectable attacks

Structural limitation: a closer look

- Fundamental structural limitations still exist
  - Optimal attacks can give unbounded impact while undetected, <u>regardless</u> of the controller and detector
  - Related to undetectable attacks



- Undetectable attacks need accurate models
  - Deviations make the attacks detectable
  - Basic principle for “Watermarking & Moving Target” defense - obfuscate part of the model

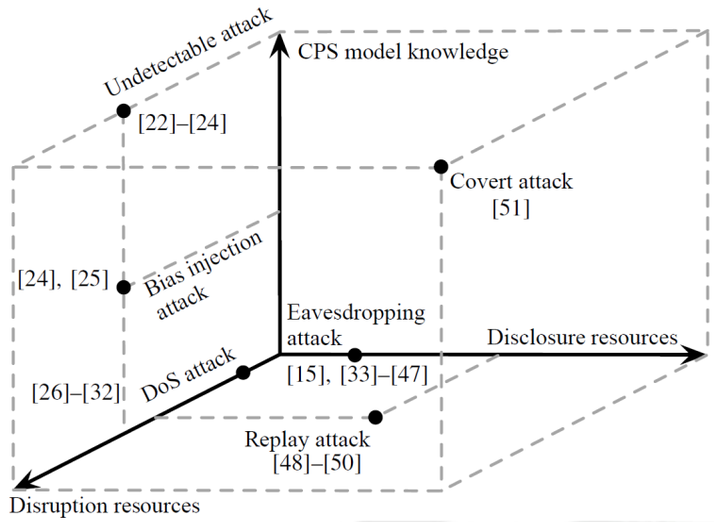


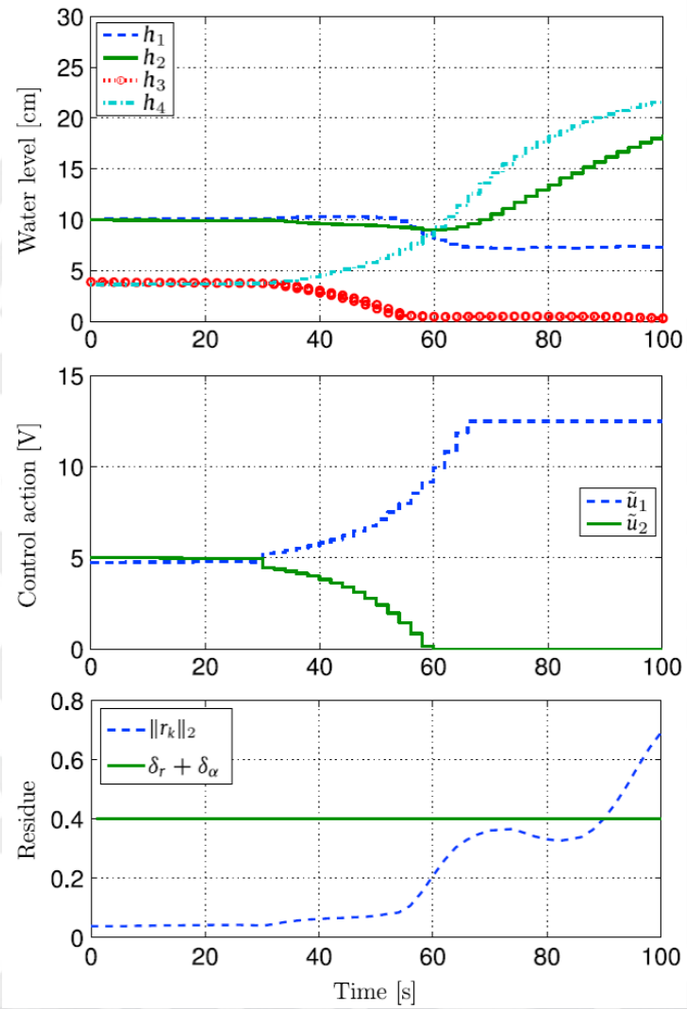
Structural limitation: a closer look

- Fundamental structural limitations still exist
  - Optimal attacks can give unbounded impact while undetected, <u>regardless</u> of the controller and detector
  - Related to undetectable attacks

- Undetectable attacks need accurate models
  - Deviations make the attacks detectable
  - Basic principle for “Watermarking & Moving Target” defense - obfuscate part of the model

- Can we capture this element in the security metrics?





Uncertainty & Robust Attacks

[Anand, under review]

Uncertainty & Robust Attacks

[Anand, under review]

- System with static parameter uncertainty

$$
\begin{aligned}
x[k+1] &= A_{cl}^{\Delta} x[k] + B_{cl}^{\Delta} a[k] \\
y_p[k] &= C_p^{\Delta} x[k] + D_p^{\Delta} a[k] \\
y_r[k] &= C_r^{\Delta} x[k] + D_r^{\Delta} a[k]
\end{aligned}
\qquad A^{\Delta} \triangleq A + \Delta A(\delta), \quad \delta \in \Omega
$$

Uncertainty & Robust Attacks

[Anand, under review]

- System with static parameter uncertainty

$$
\begin{aligned}
x[k+1] &= A_{cl}^{\Delta} x[k] + B_{cl}^{\Delta} a[k] \\
y_p[k] &= C_p^{\Delta} x[k] + D_p^{\Delta} a[k] \\
y_r[k] &= C_r^{\Delta} x[k] + D_r^{\Delta} a[k]
\end{aligned}
\qquad A^{\Delta} \triangleq A + \Delta A(\delta), \quad \delta \in \Omega
$$

- Robust attack policy

$$
\begin{aligned}
\sup_{a \in \ell_{2e}} \quad & \mathbb{E}_{\Omega}\{||y_p(\delta)||_{\ell_2}^2\} \\
\text{s.t.} \quad & ||y_r(\delta)||_{\ell_2}^2 \leq 1, \quad \forall \delta \in \Omega \\
& x(\delta)[0] = 0.
\end{aligned}
$$

Uncertainty & Robust Attacks

[Anand, under review]

- System with static parameter uncertainty

$$
\begin{aligned}
x[k+1] &= A_{cl}^{\Delta} x[k] + B_{cl}^{\Delta} a[k] \\
y_p[k] &= C_p^{\Delta} x[k] + D_p^{\Delta} a[k] \\
y_r[k] &= C_r^{\Delta} x[k] + D_r^{\Delta} a[k]
\end{aligned}
\qquad A^{\Delta} \triangleq A + \Delta A(\delta), \quad \delta \in \Omega
$$

- Robust attack policy

$$
\begin{aligned}
\sup_{a \in \ell_{2e}} \quad & \mathbb{E}_{\Omega}\{||y_p(\delta)||_{\ell_2}^2\} \\
\text{s.t.} \quad & ||y_r(\delta)||_{\ell_2}^2 \leq 1, \quad \forall \delta \in \Omega \\
& x(\delta)[0] = 0.
\end{aligned}
$$

- Sampled approximation

$$
\begin{aligned}
\sup_{a \in \ell_{2e}} \quad & \frac{1}{N_s} \sum_{i=1}^{N_s} ||y_p(\delta_i)||_{\ell_2}^2\} \\
\text{s.t.} \quad & ||y_r(\delta_i)||_{\ell_2}^2 \leq 1, \quad \forall i \in \{1, \ldots, N_s\} \\
& x(\delta_i)[0] = 0.
\end{aligned}
$$

Example

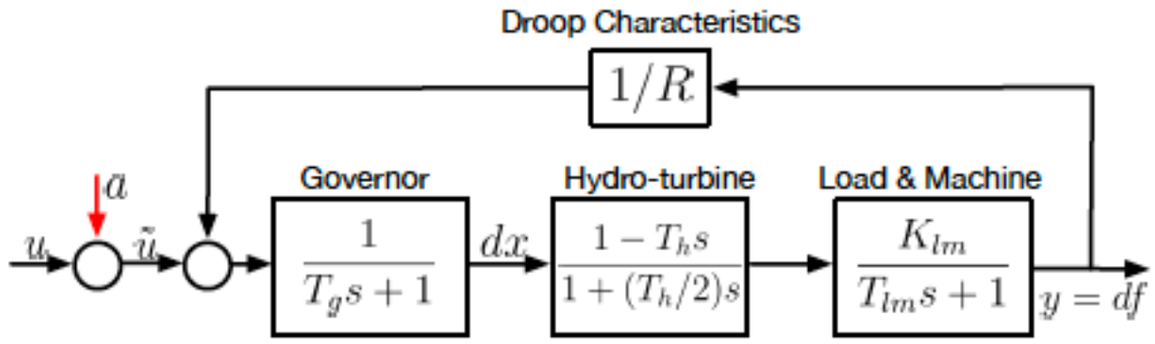


Fig. 4. Power generating system with a hydro turbine.

SYSTEM PARAMETERS

K_{lm}	1	T_{lm}	6
T_g	0.2	R	0.05
T_h	[4 6]	T_s	0.1

Example

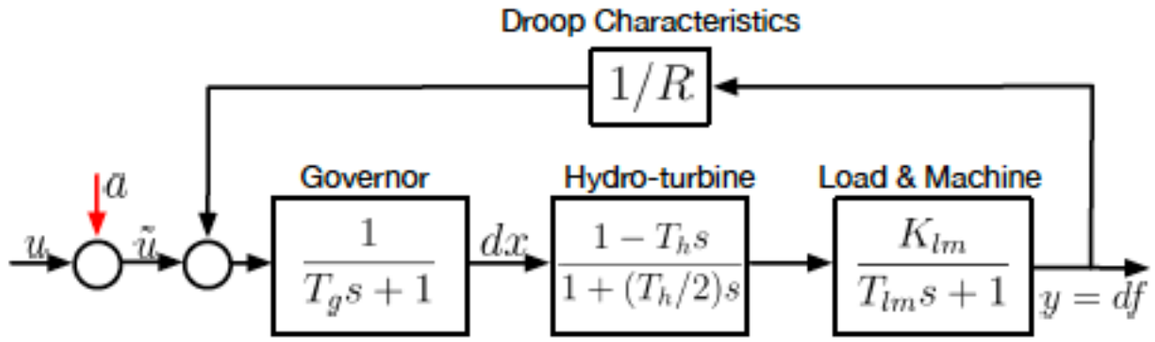


Fig. 4. Power generating system with a hydro turbine.

SYSTEM PARAMETERS

K_{lm}	1	T_{lm}	6
T_g	0.2	R	0.05
T_h	[4 6]	T_s	0.1

Without uncertainty:

- Unbounded impact for any parameter value

Example

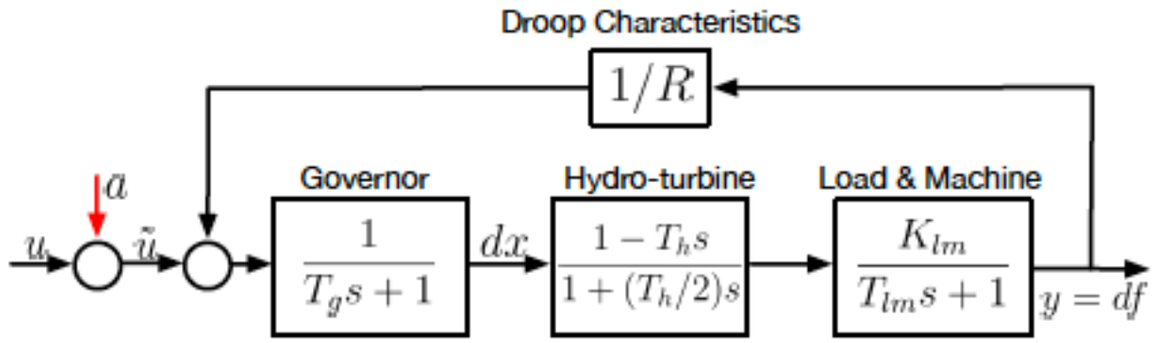


Fig. 4. Power generating system with a hydro turbine.

SYSTEM PARAMETERS

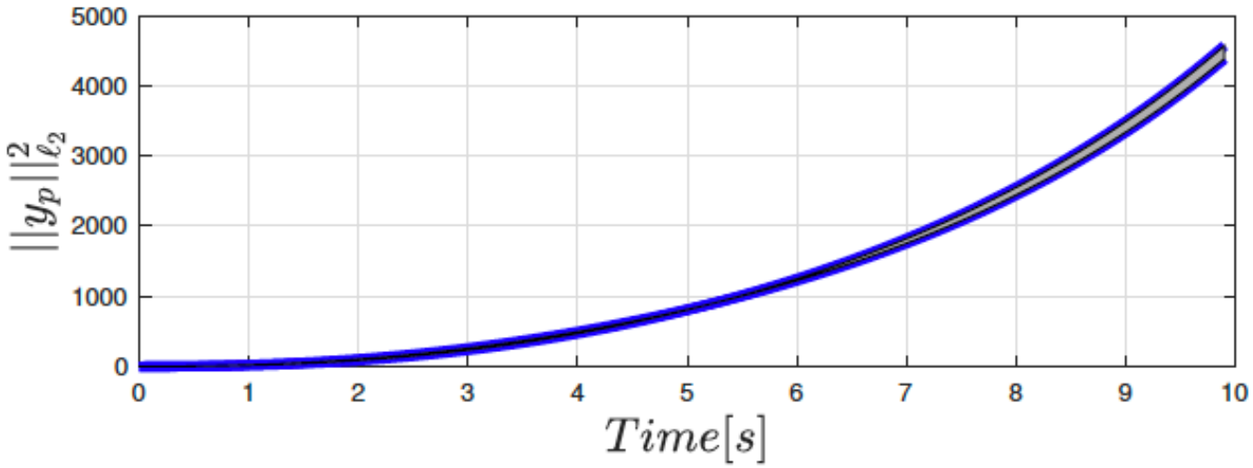
K_{lm}	1	T_{lm}	6
T_g	0.2	R	0.05
T_h	[4 6]	T_s	0.1

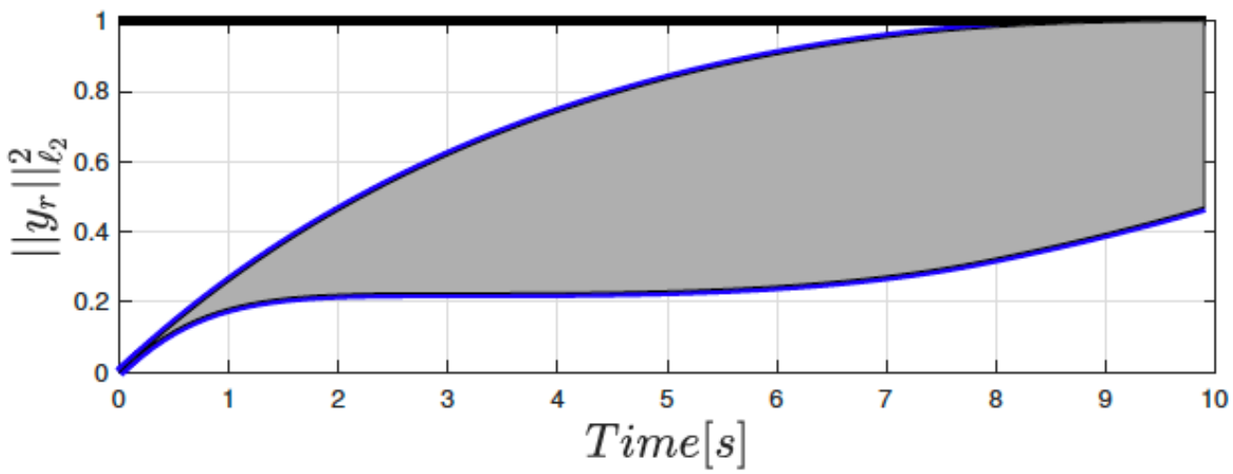
Without uncertainty:

- Unbounded impact for any parameter value

With uncertain T_h:

- Impact becomes bounded when T_h is uncertain & attack is robust





Example

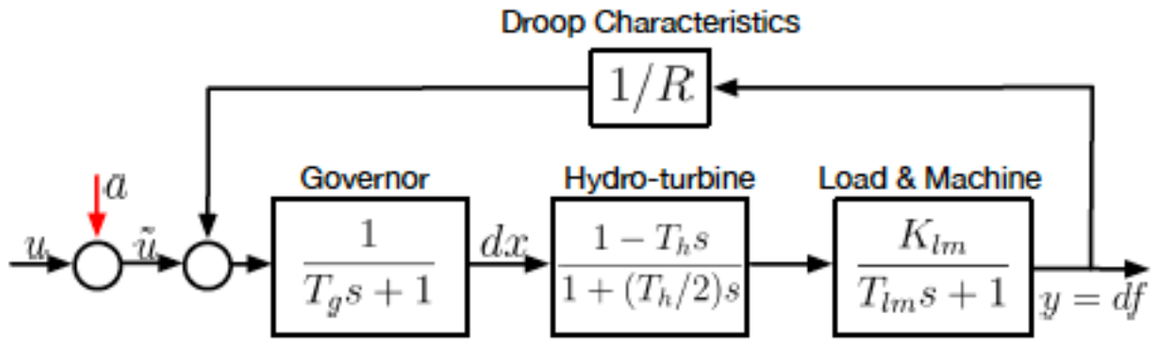


Fig. 4. Power generating system with a hydro turbine.

SYSTEM PARAMETERS

K_{lm}	1	T_{lm}	6
T_g	0.2	R	0.05
T_h	[4 6]	T_s	0.1

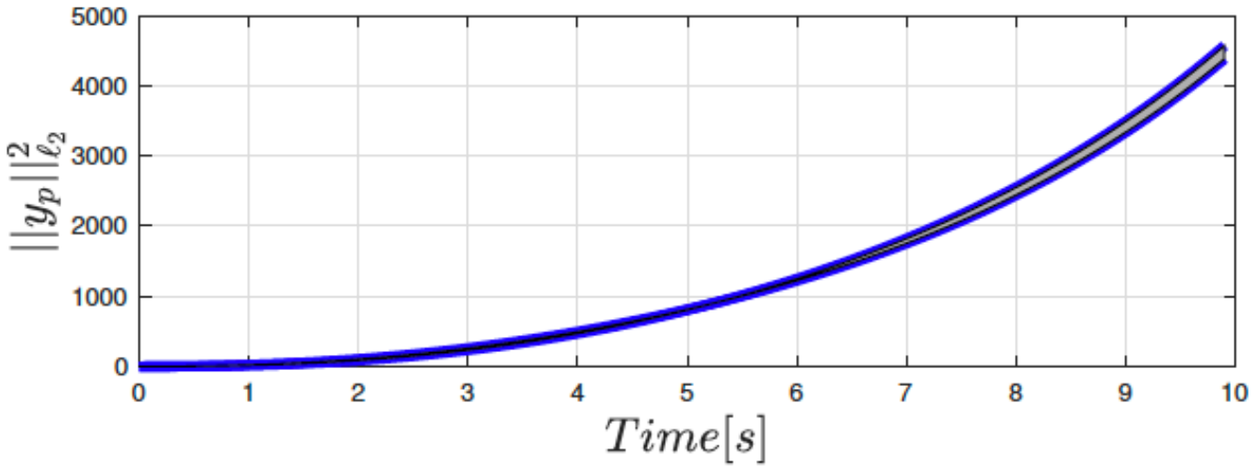
Without uncertainty:

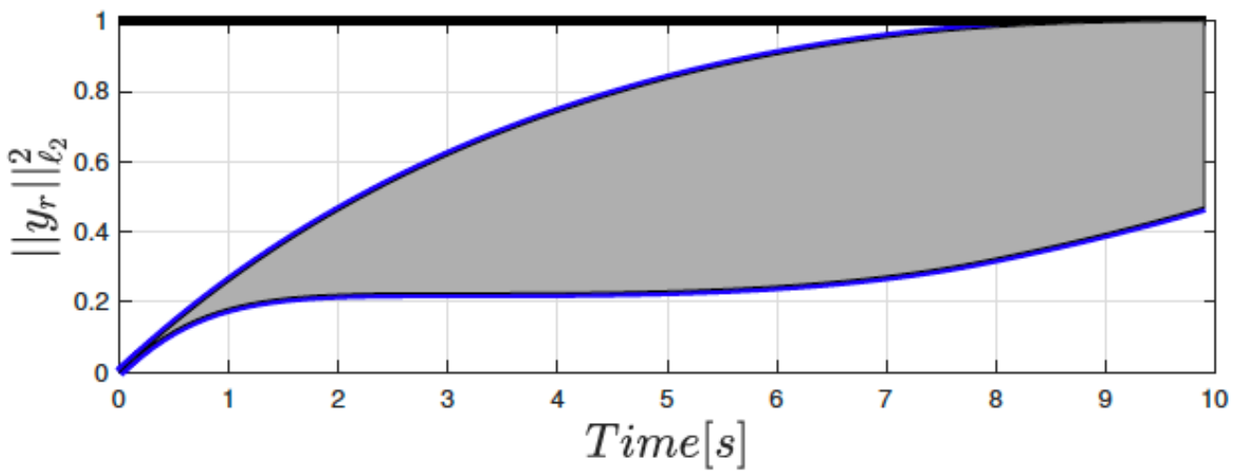
- Unbounded impact for any parameter value

With uncertain T_h:

- Impact becomes bounded when T_h is uncertain & attack is robust

- "Uncertainty as a defense" can be incorporated by design
  - Watermarking, moving target, weak encryption, …





UPPSALA UNIVERSITET

Summary

UPPSALA UNIVERSITET

Summary

- Secure Control Systems - a risk management perspective
  - Example: undetectable attacks

UPPSALA UNIVERSITET

Summary

- Secure Control Systems - a risk management perspective
  - Example: undetectable attacks

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

Summary

- Secure Control Systems - a risk management perspective
  - Example: undetectable attacks

- 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
  - **Analysis** based on LMIs
  - Uncertainty at the adversary can also be incorporated
  - Variations of adversary models and their metrics

Summary

- Secure Control Systems - a risk management perspective
  - Example: undetectable attacks

- 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
  - **Analysis** based on LMIs
  - Uncertainty at the adversary can also be incorporated
  - Variations of adversary models and their metrics

- Novel security metrics enables new defense mechanisms
  - Controller / Detector **design** through BMIs
  - “Uncertainty as defense”