Risk Management and Game Theory for Securing Control Systems

The Subtle Interplays between Adversary Models, Security Risk Metrics, and Uncertainty

André Teixeira

Associate Professor

Dept. of Information Technology

Uppsala University

Uppsala Secure Learning and Control Lab

Swedish Research Council

Swedish Foundation for Strategic Research

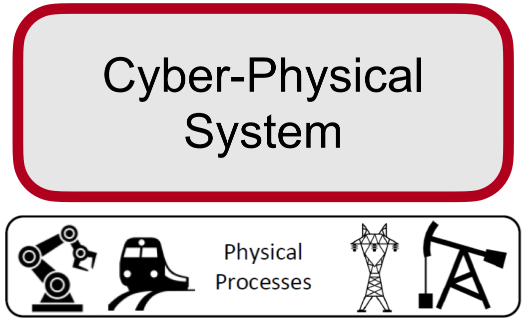
IEEE CDC, December 12, 2023

Uppsala Universitet

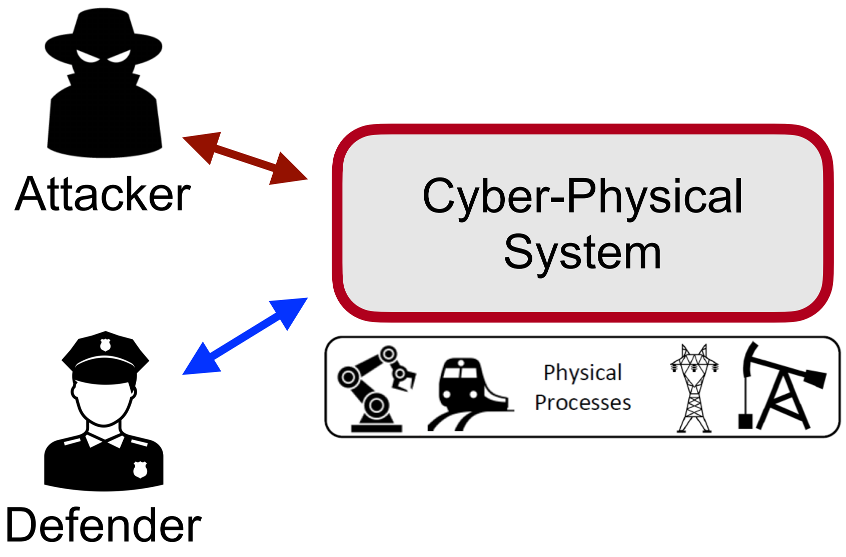
Outline

- Security Risk Management
- Scenario and Threat Models
- Security Metrics and Game-Theoretic Design
- Security under Model Uncertainty
- Probabilistic Risk Measures and Game-Theoretic Design
- Conclusions and Remarks

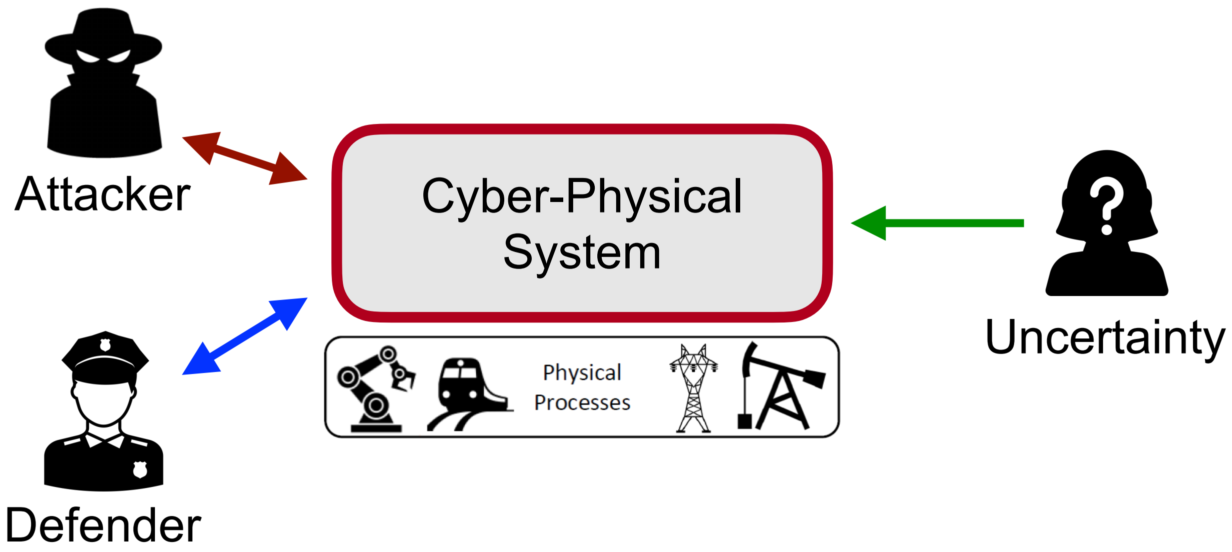
“The Security Game”: key ingredients



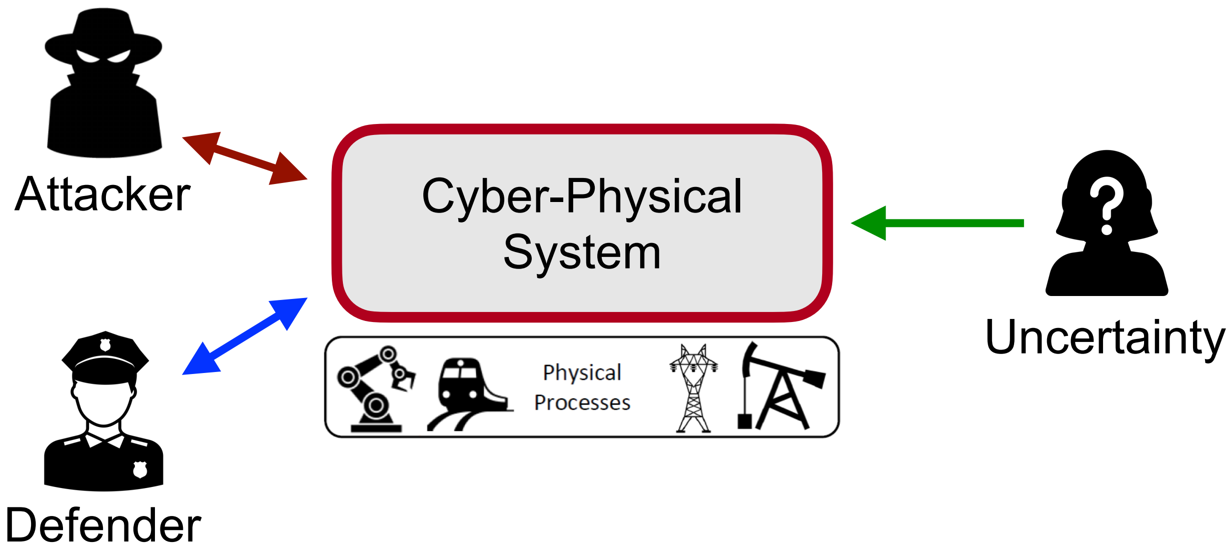
“The Security Game”: key ingredients



“The Security Game”: key ingredients



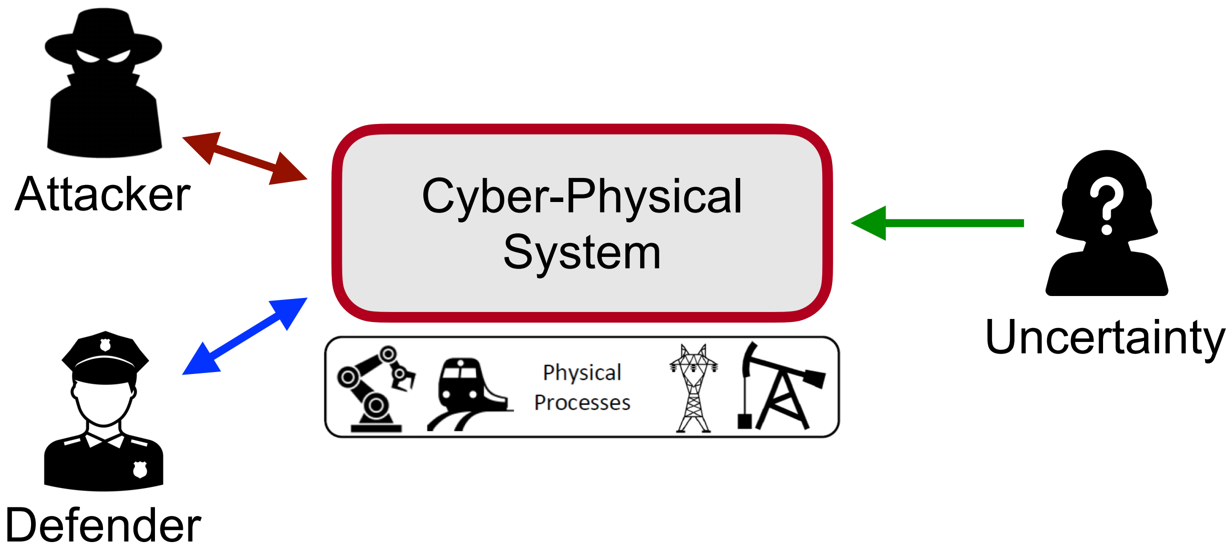
“The Security Game”: key ingredients



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:

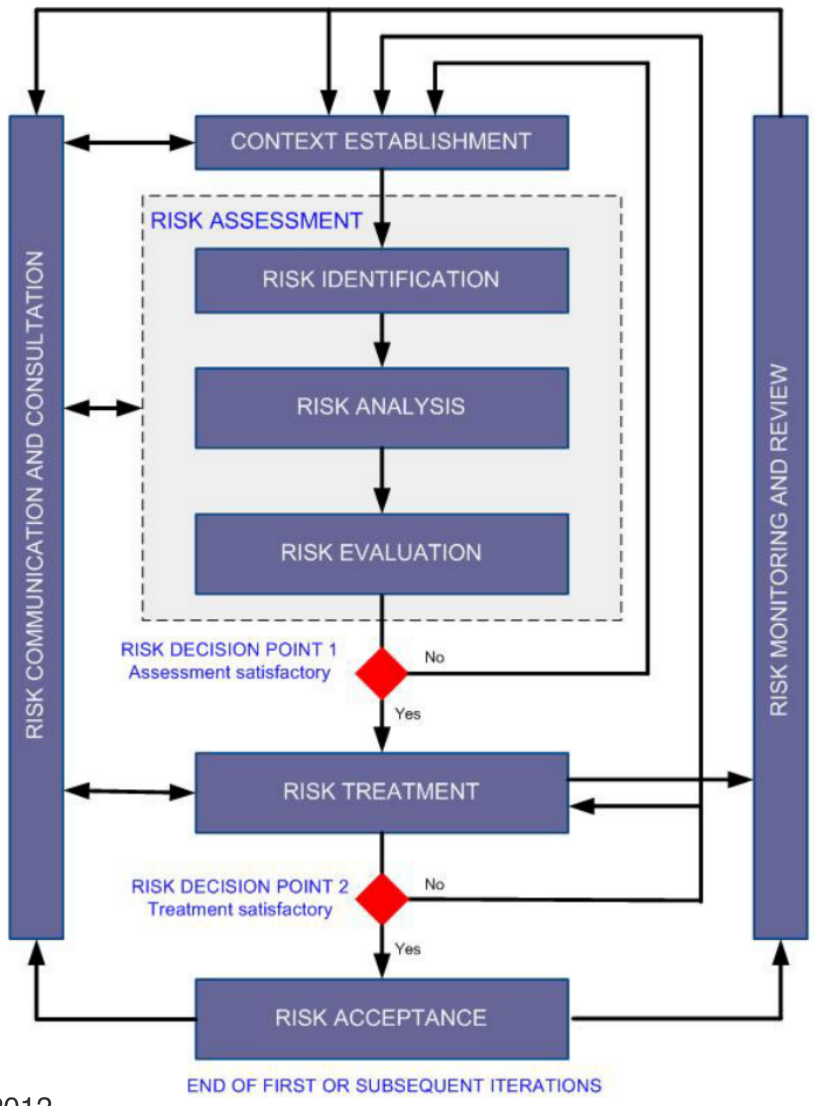
- **Which threats** should we care about?
- **What impact** can we expect from attacks?
- **Which resources** should we **protect**, and how?

- How to find answers: **Risk Management & Game Theory + Control Theory + Statistical Learning**

Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]



[ISO 31000]

S. Sridhar, et al., "Cyber–Physical System Security for the Electric Power Grid," in *Proceedings of the IEEE*, 2012

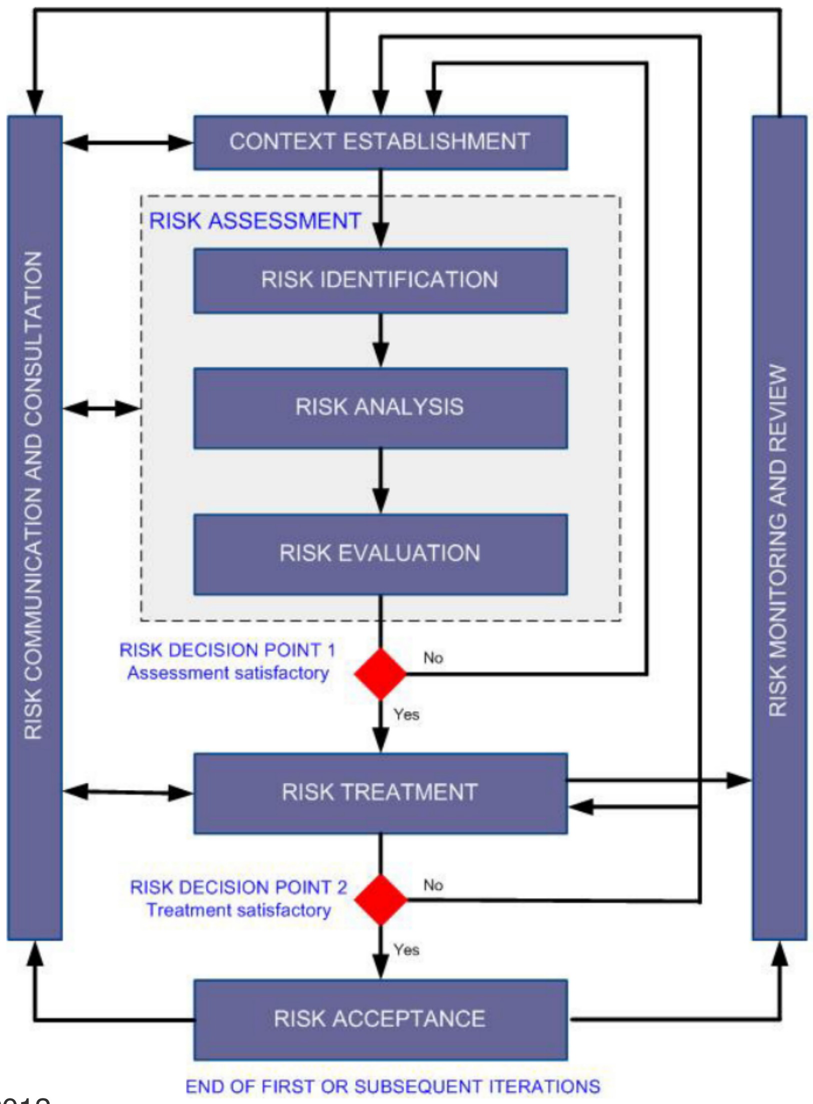
Milosevic et al. "Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity". Int. Journal of Robust and Nonlinear Control 2020.

Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management



[ISO 31000]

S. Sridhar, et al., "Cyber–Physical System Security for the Electric Power Grid," in *Proceedings of the IEEE*, 2012

Milosevic et al. "Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity". Int. Journal of Robust and Nonlinear Control 2020.

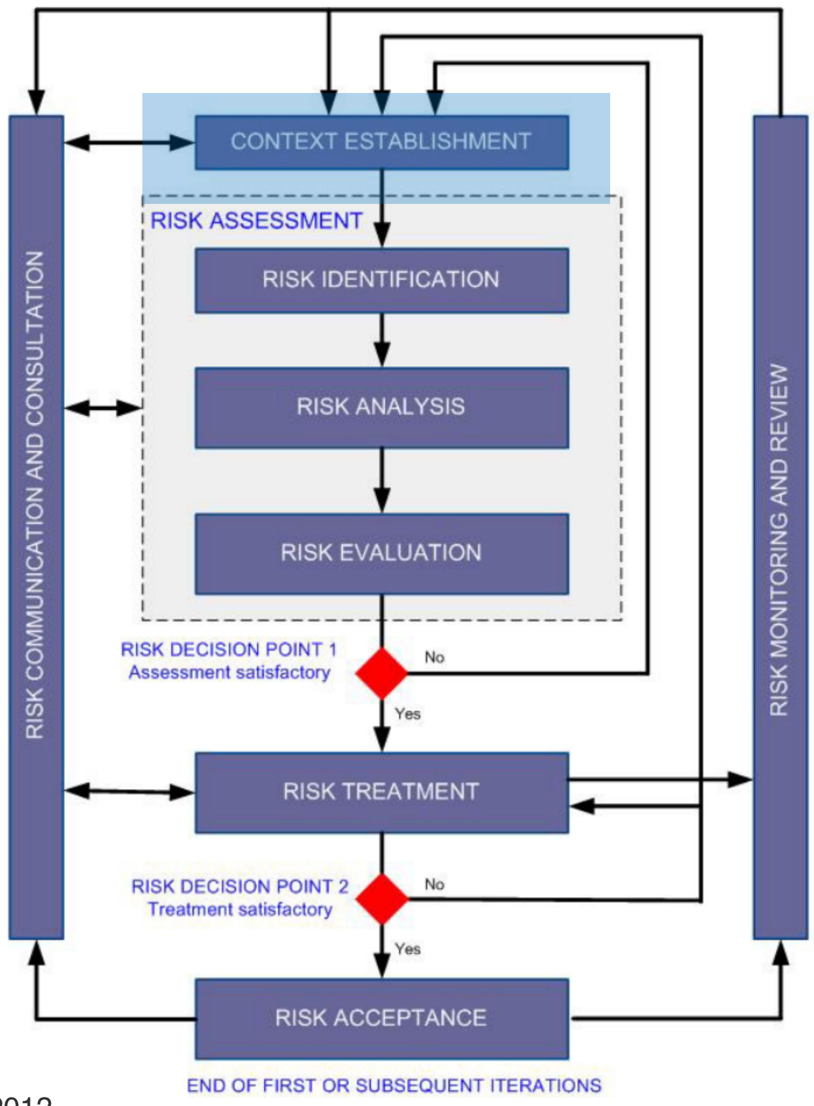
Risk Management Cycle

Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

- Scenario characterization
    - Models, Scenarios, Objectives



[ISO 31000]

S. Sridhar, et al., "Cyber–Physical System Security for the Electric Power Grid," in *Proceedings of the IEEE*, 2012

Milosevic et al. “Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity”. Int. Journal of Robust and Nonlinear Control 2020.

Risk Management Cycle

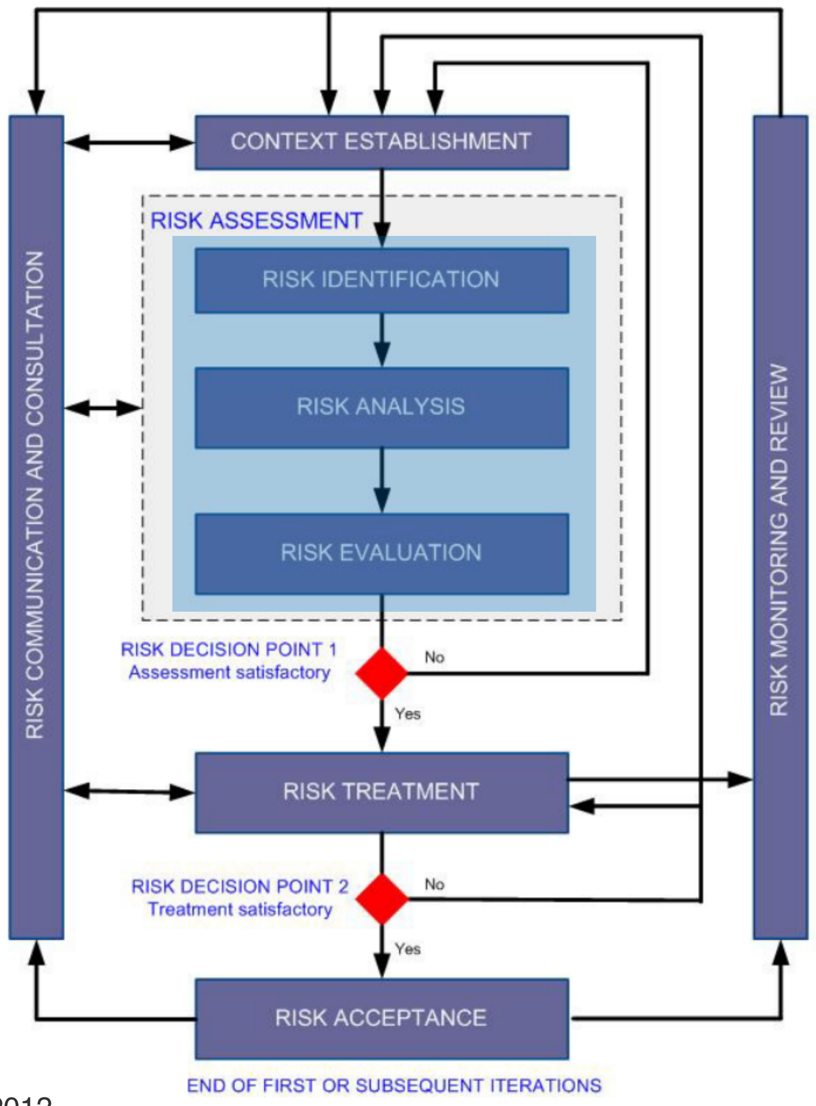
Risk = (Scenario, Likelihood, Impact)

[Kaplan & Garrick, 1981]

Main steps in risk management

- Scenario characterization
    - Models, Scenarios, Objectives

- Risk Analysis
    - Likelihood Assessment
    - Impact Assessment



[ISO 31000]

S. Sridhar, et al., "Cyber–Physical System Security for the Electric Power Grid," in *Proceedings of the IEEE*, 2012

Milosevic et al. “Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity”. Int. Journal of Robust and Nonlinear Control 2020.

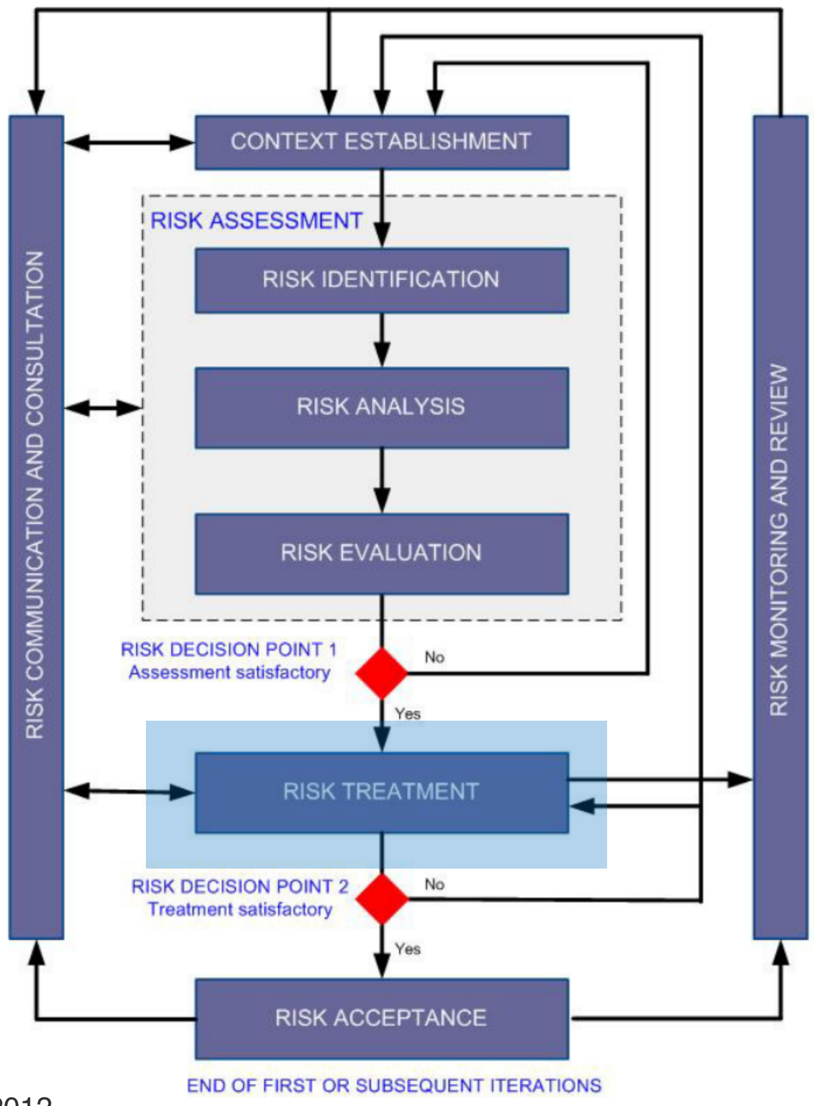
Risk Management Cycle

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]

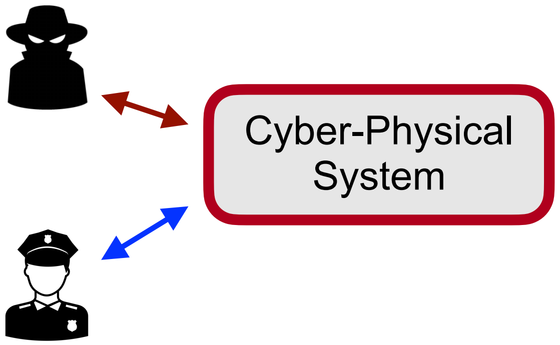
S. Sridhar, et al., "Cyber–Physical System Security for the Electric Power Grid," in *Proceedings of the IEEE*, 2012

Milosevic et al. "Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity". Int. Journal of Robust and Nonlinear Control 2020.

Outline

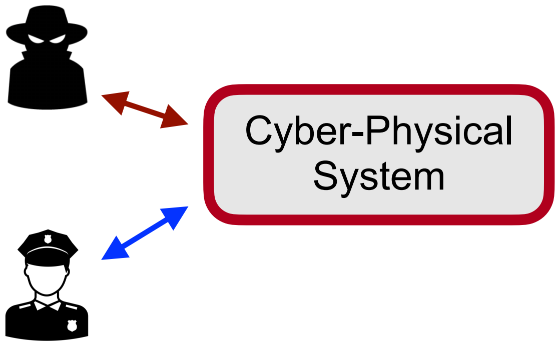
- Security Risk Management
- **Scenario and Threat Models**
- Security Metrics and Game-Theoretic Design
- Security under Model Uncertainty
- Probabilistic Risk Measures and Game-Theoretic Design
- Conclusions and Remarks

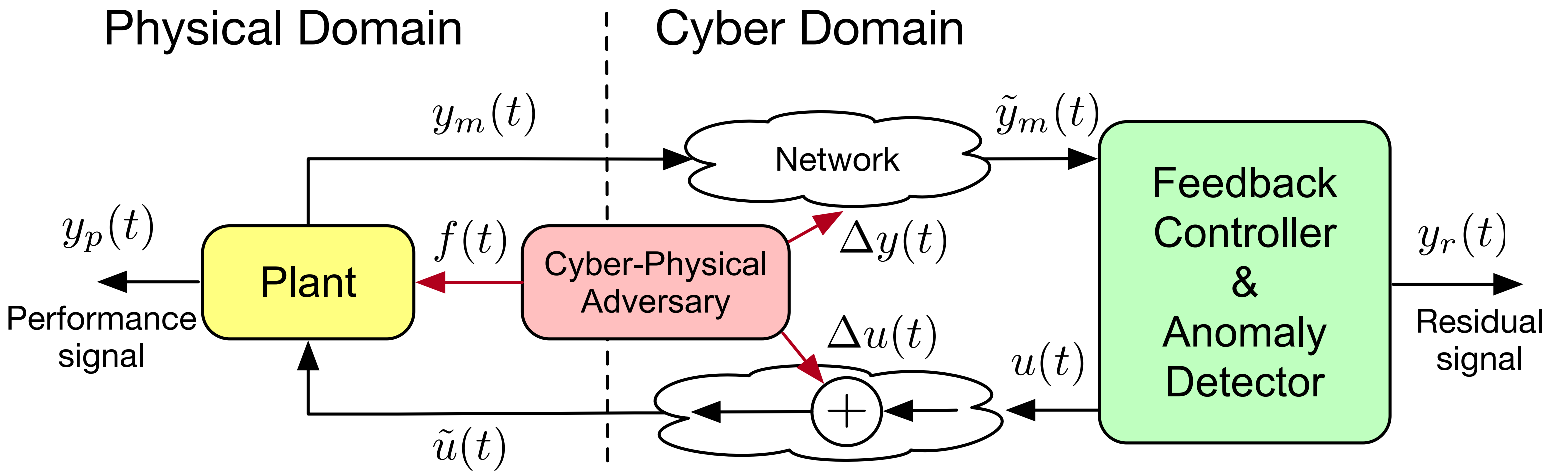
System Model



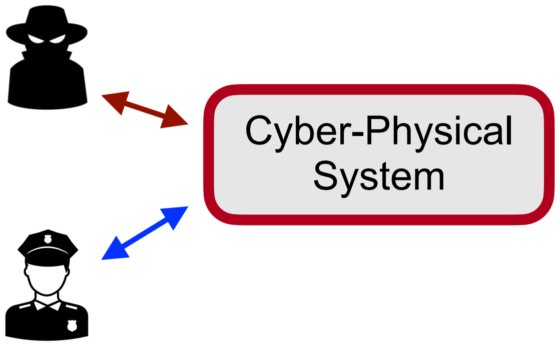
UPPSALA
UNIVERSITET

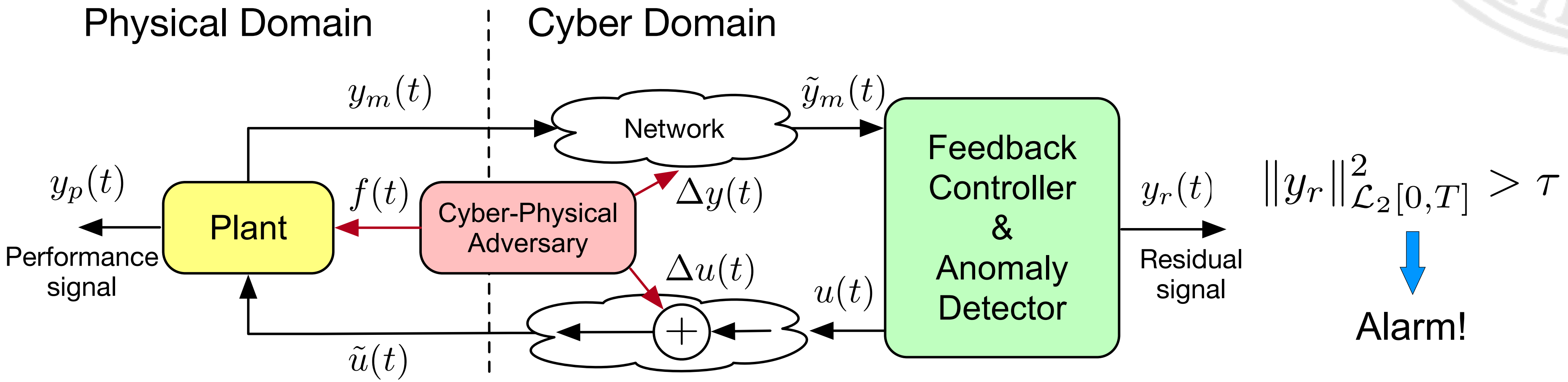
System Model





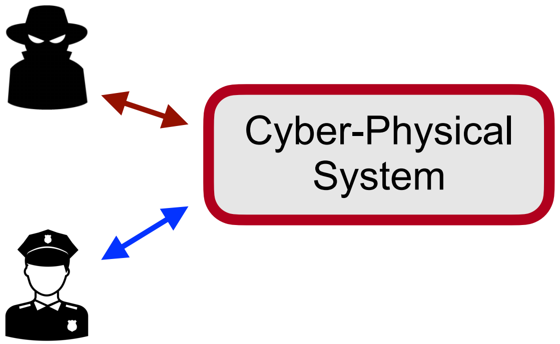
System Model

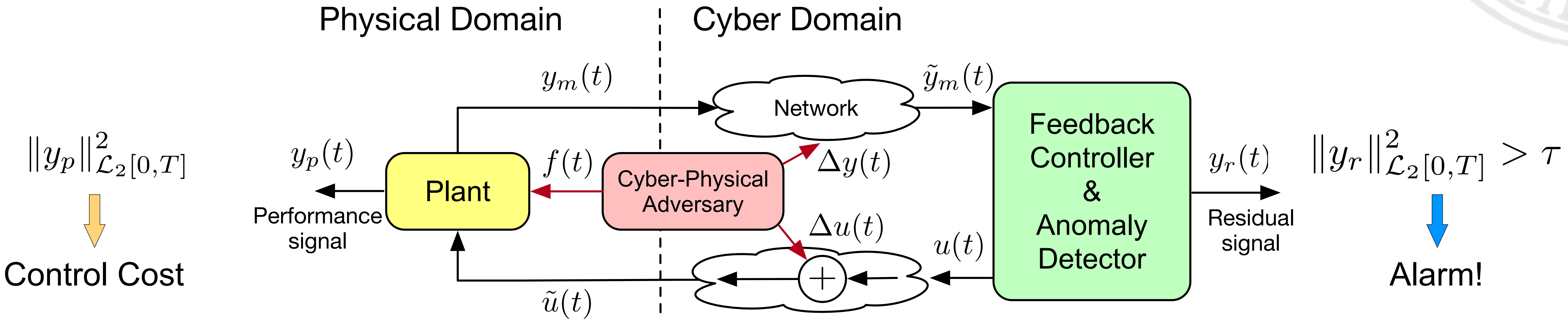




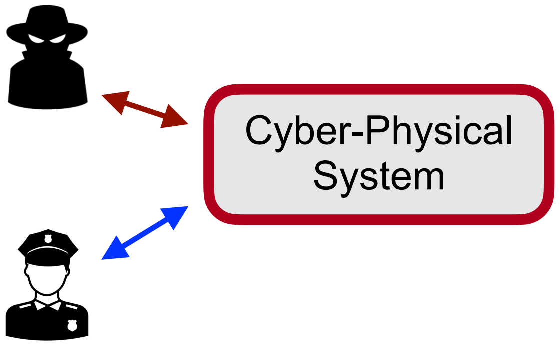
UPPSALA UNIVERSITET

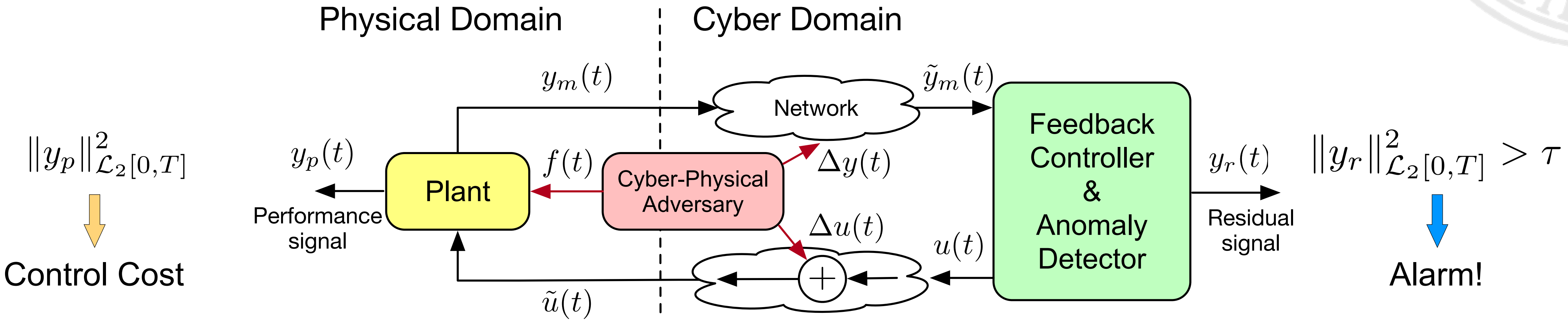
System Model





System Model





$y_p(t)$

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

$a(t)$

Attack signal

$y_r(t)$

Closed-loop system

Adversary Models

Adversary Models

Key elements [Do 2019]
- Goals
- Assumptions
- Capabilities

Do et al., "The role of the adversary model in applied security research", Computers & Security, 2019

Adversary Models

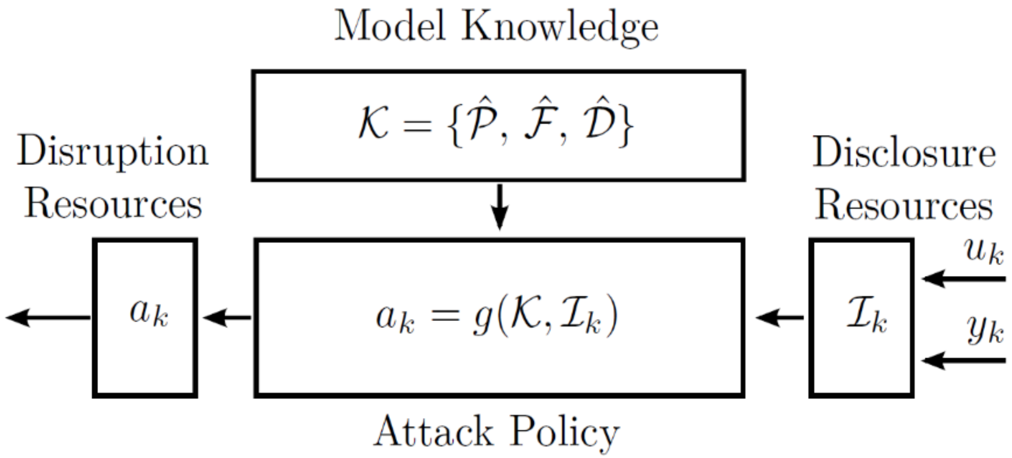
Key elements [Do 2019]

- Goals
- Assumptions
- Capabilities

⚠ Adversary models are extremely important to define security!
They define *what* the system is (in)secure against.

Do et al., "The role of the adversary model in applied security research", Computers & Security, 2019

Adversary Models



Key elements [Do 2019]
- Goals
- Assumptions
- Capabilities

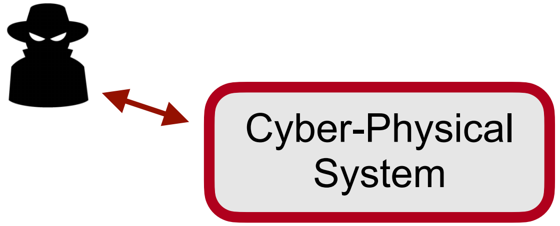
Adversary models are extremely important to define security!
They define *what* the system is (in)secure against.

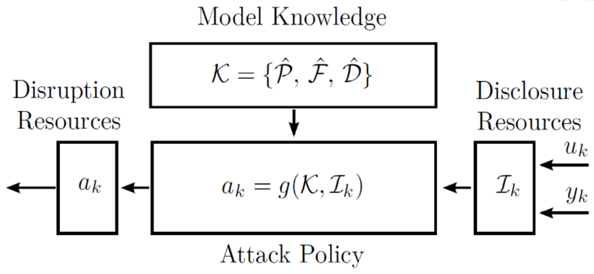
- **Attack policy:** Goal of the attack? Destroy equipment, increase costs, *remain undetected*…
- **CPS model knowledge:** Adversary knows models of plant and controller? Better models increase possibility for stealthy attacks…
- **Disruption/disclosure resources:** Which channels can the adversary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

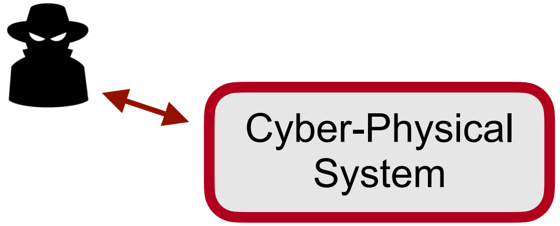
[Teixeira *et al*., Automatica, 2015]

Do et al., "The role of the adversary model in applied security research", Computers & Security, 2019

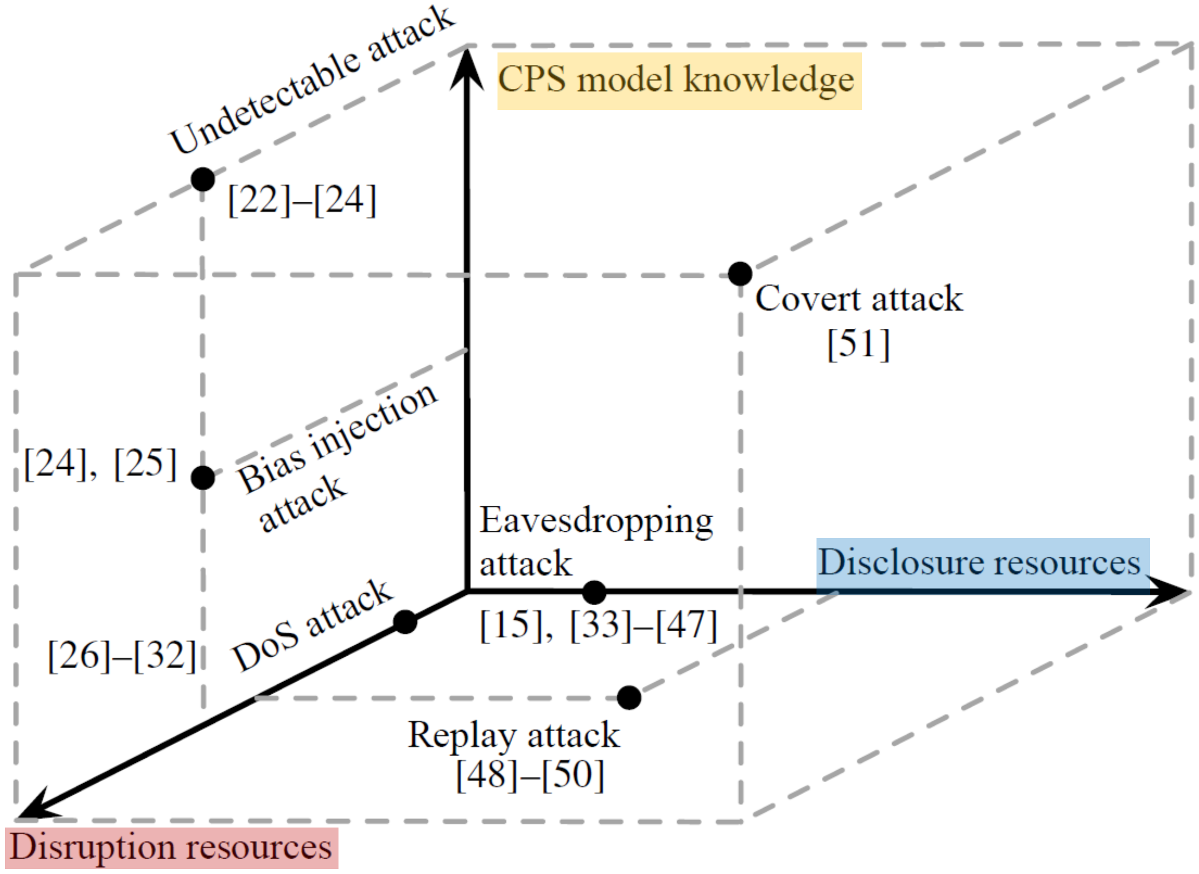
Attack Scenarios

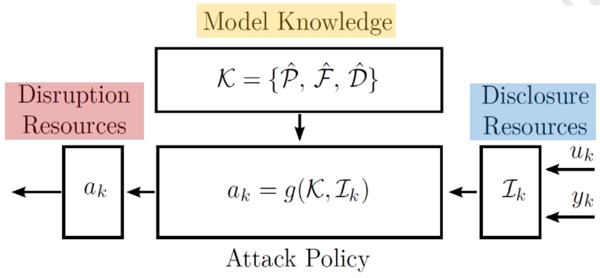




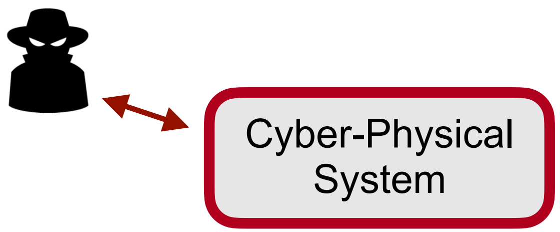


Attack Scenarios

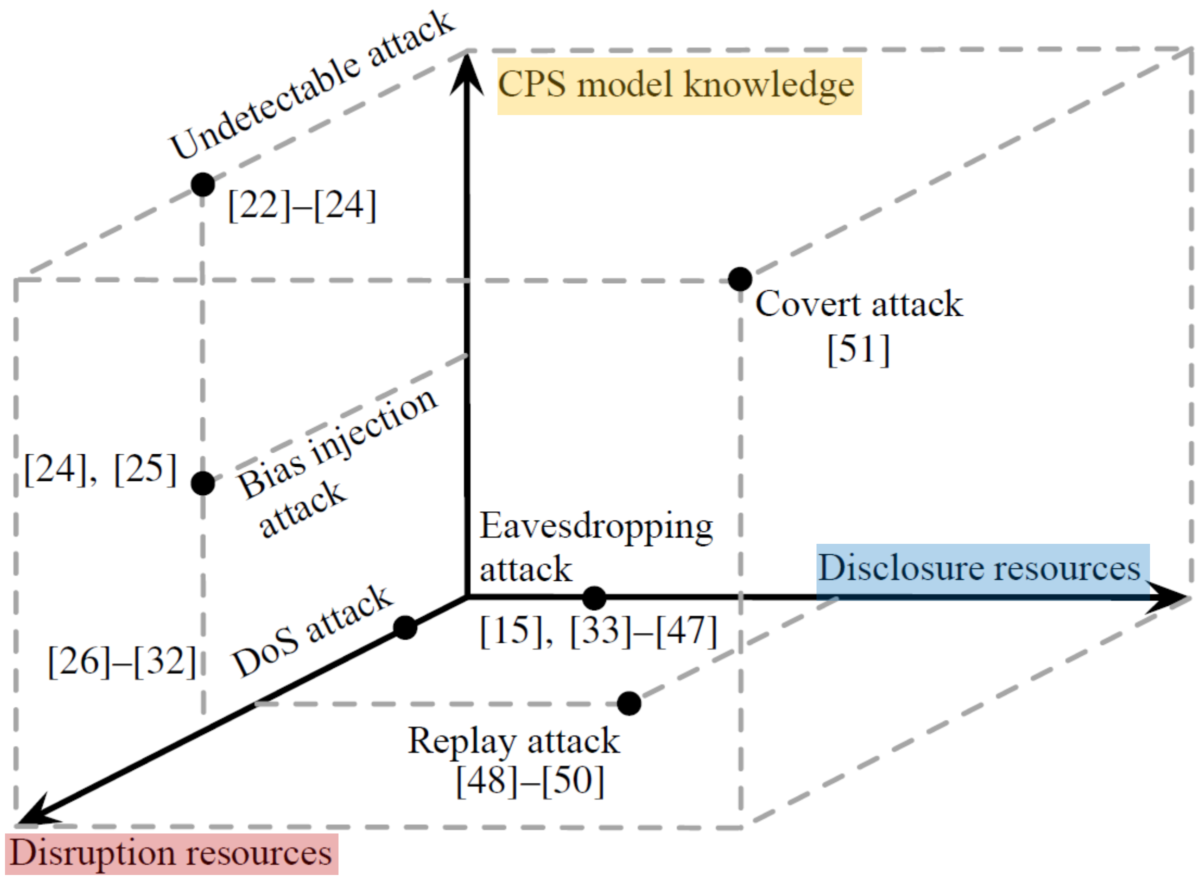


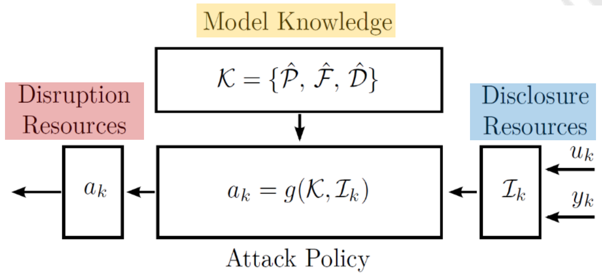


[Chong *et al.*, ECC, 2019] [Teixeira *et al.*, Automatica, 2015]



Attack Scenarios





Adversary Models

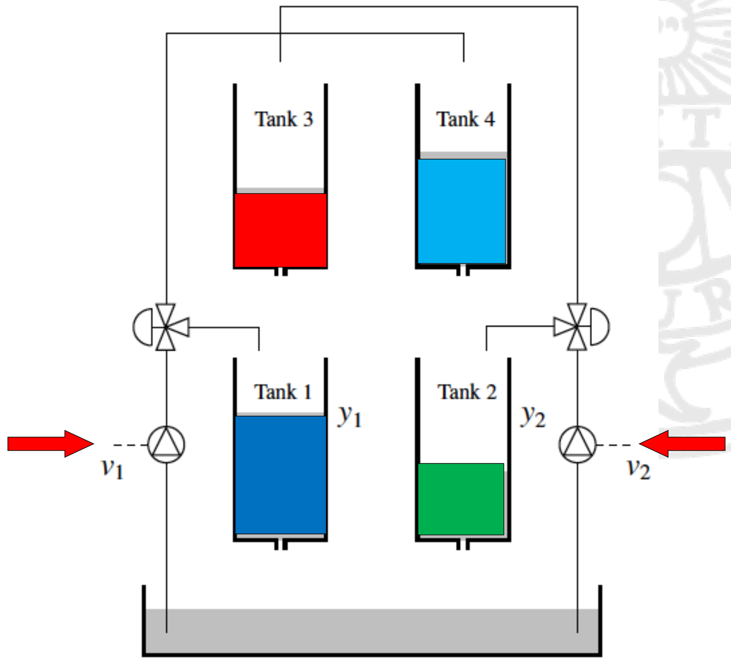
- How does the adversary behave against the system?

Security Analysis:

- Can it evade detection?
- Can it violate safety?
- How complex/likely is it?

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

Example: Zero Dynamics Attack



Example: Zero Dynamics Attack

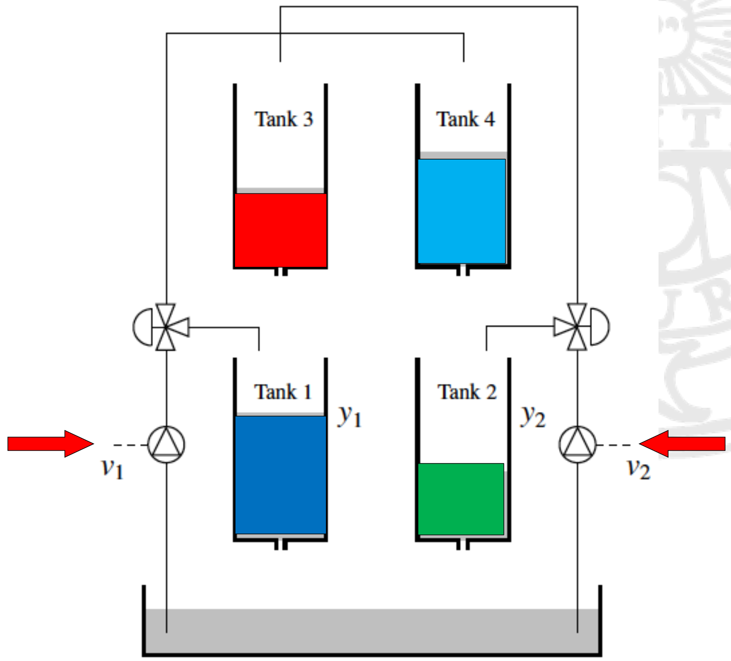
$y_p(t)$

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

$a(t)$

$y_r(t)$

Closed-loop system



Example: Zero Dynamics Attack

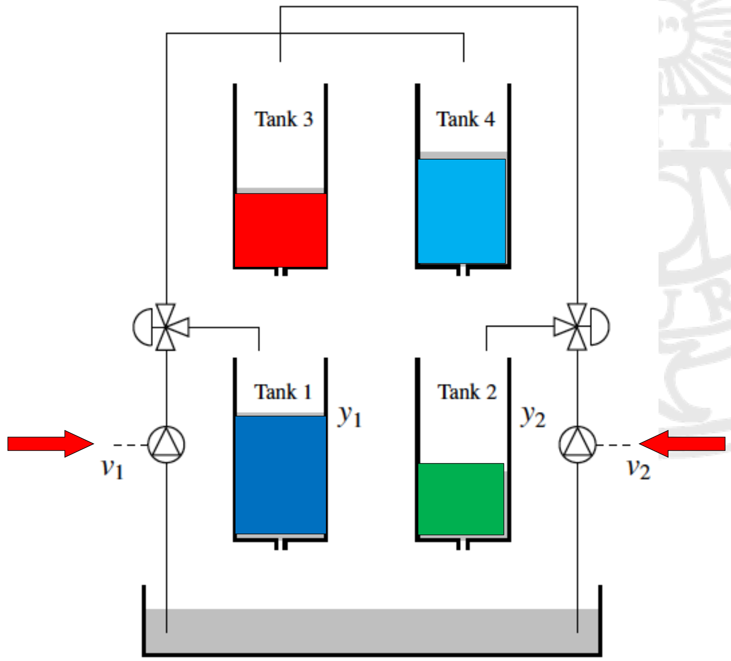
$y_p(t)$

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

$a(t)$

$y_r(t)$

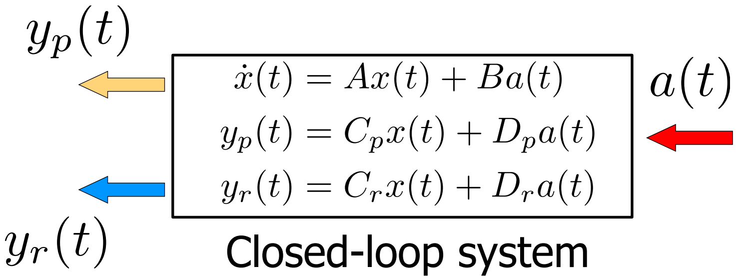
Closed-loop system

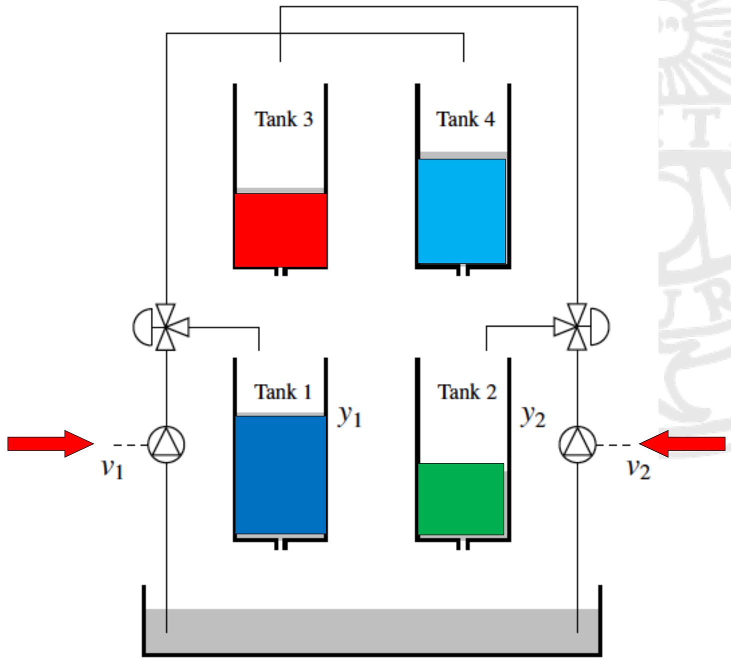


- (Discrete-time) zero dynamics characterized by:

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

Example: Zero Dynamics Attack





- (Discrete-time) zero dynamics characterized by:

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

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

Example: Zero Dynamics Attack

$y_p(t)$

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

$a(t)$

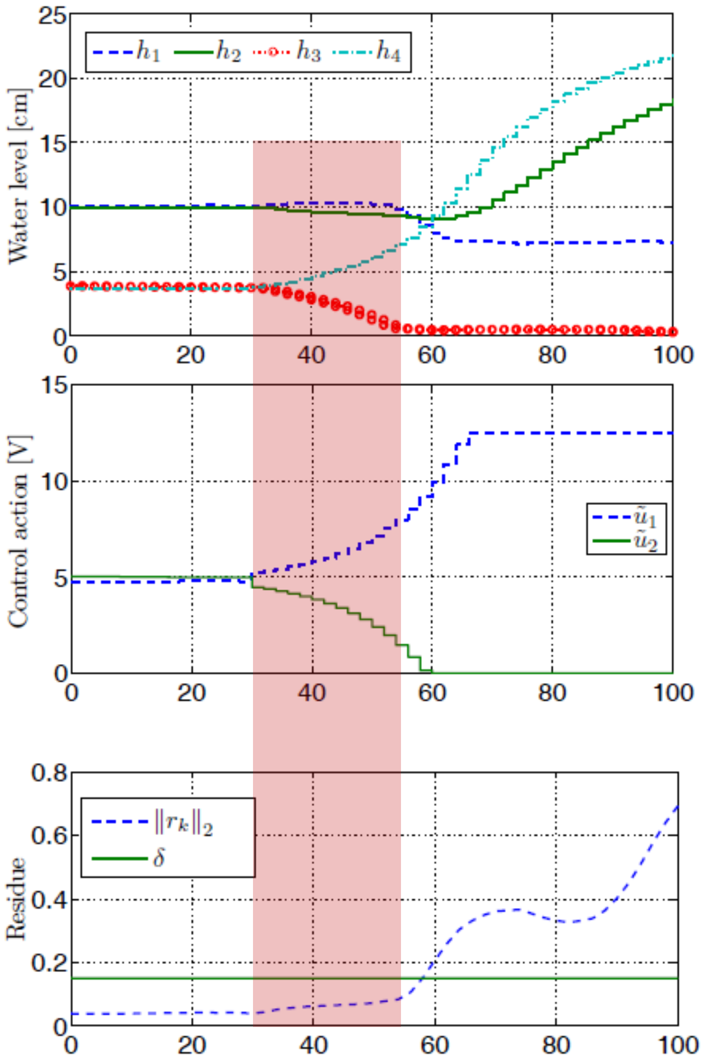
$y_r(t)$

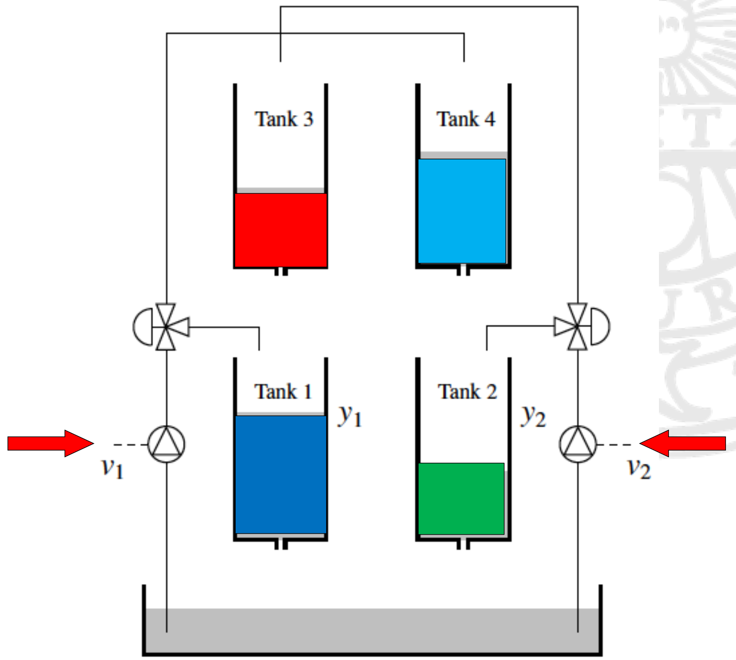
Closed-loop system

- (Discrete-time) zero dynamics characterized by:

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

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





UPPSALA UNIVERSITET

Example: Zero Dynamics Attack

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

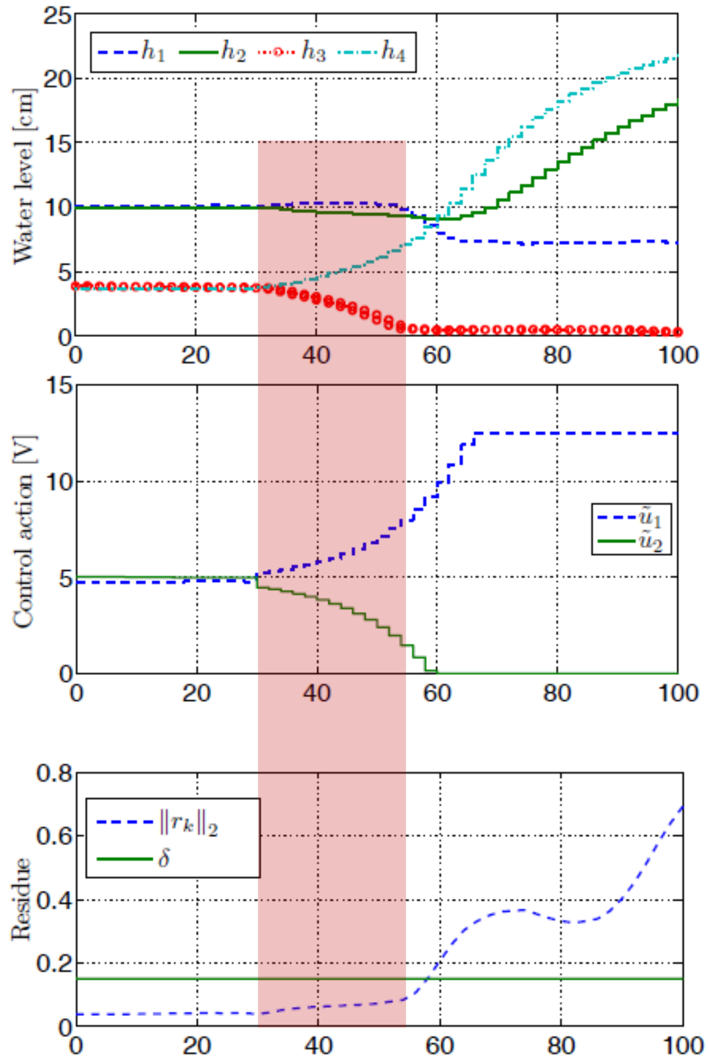
$y_r(t)$ ←

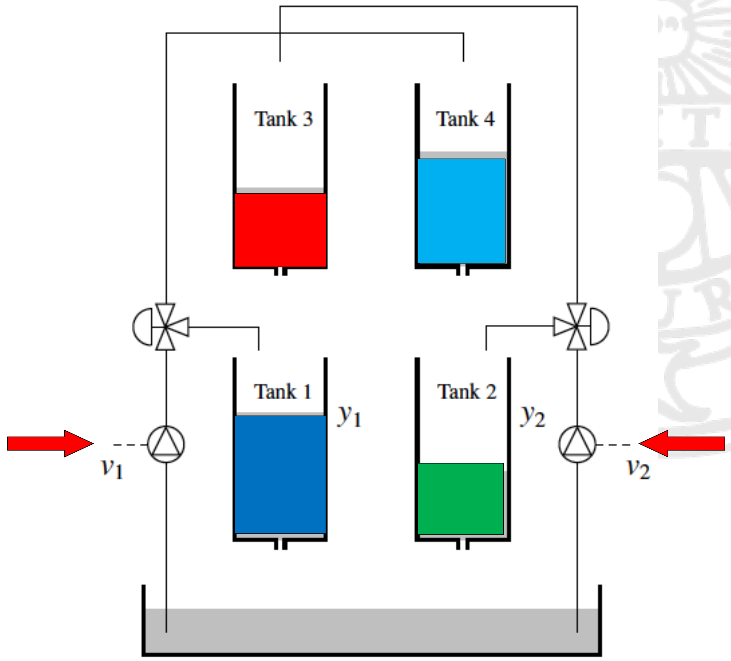
Closed-loop system

- (Discrete-time) zero dynamics characterized by:

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

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





Attack is undetected during the "linear" regime.

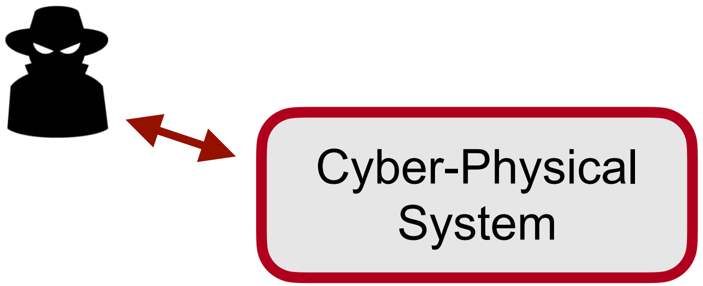
Attack impact is significant: empties Tank 3.

UPPSALA UNIVERSITET

Defense: Active Detection of Attacks

- Attack characterized by:

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

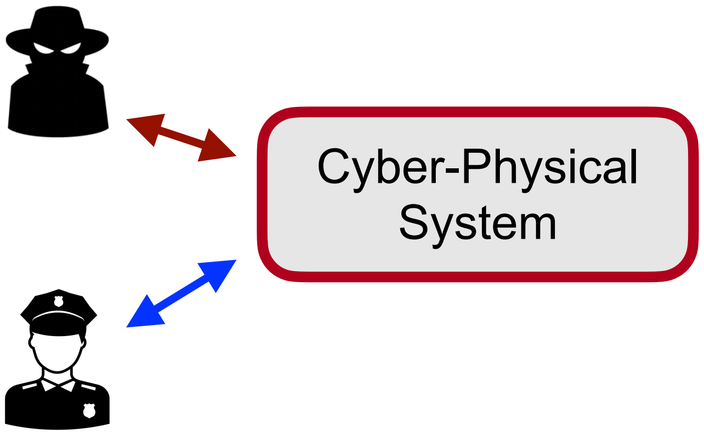


Defense: Active Detection of Attacks

- Attack characterized by:

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

- Moving Target Defense (MTD):
  - Modify the system dynamics



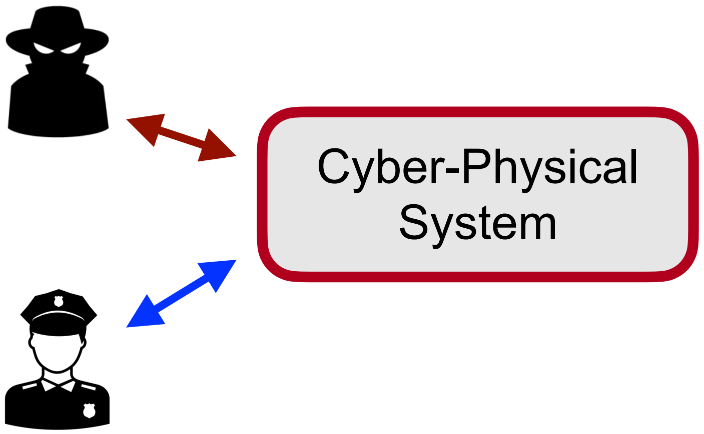
MTD creates uncertainty in the adversary.

Defense: Active Detection of Attacks

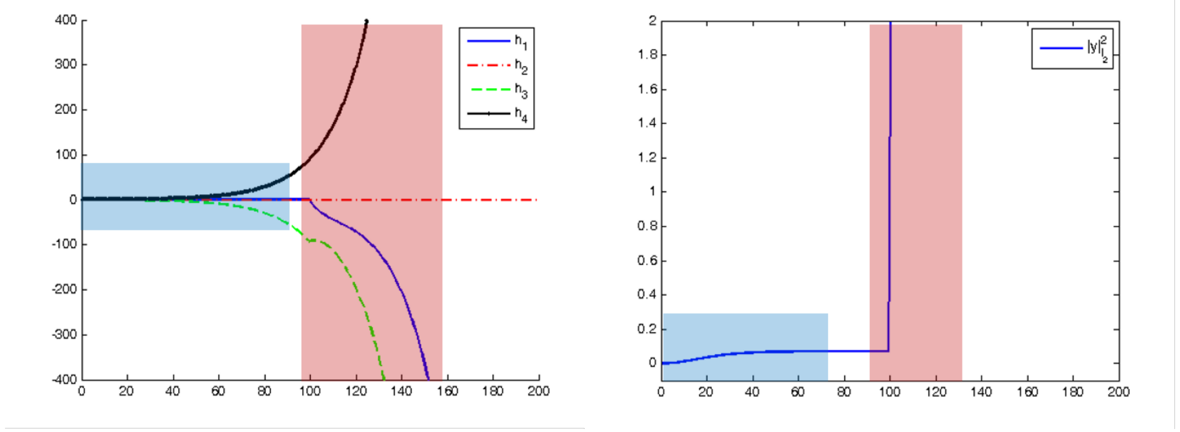
- Attack characterized by:

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

- Moving Target Defense (MTD):
  - Modify the system dynamics



Example: connect Tank 1 to Tank 3



Teixeira et al, “Revealing Stealthy Attacks in Control Systems”. Allerton 2012

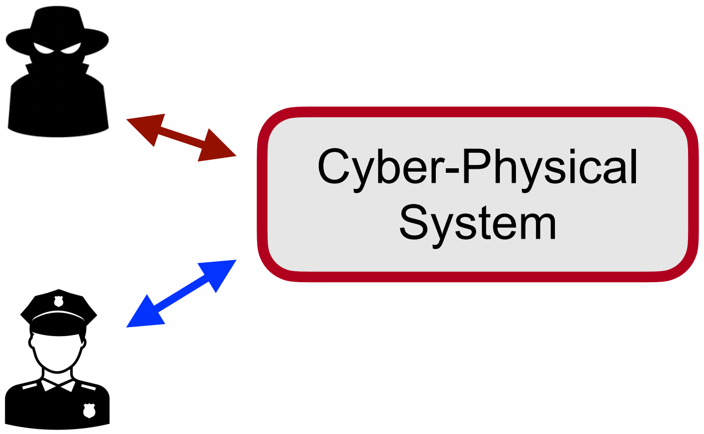
MTD creates uncertainty in the adversary.

Defense: Active Detection of Attacks

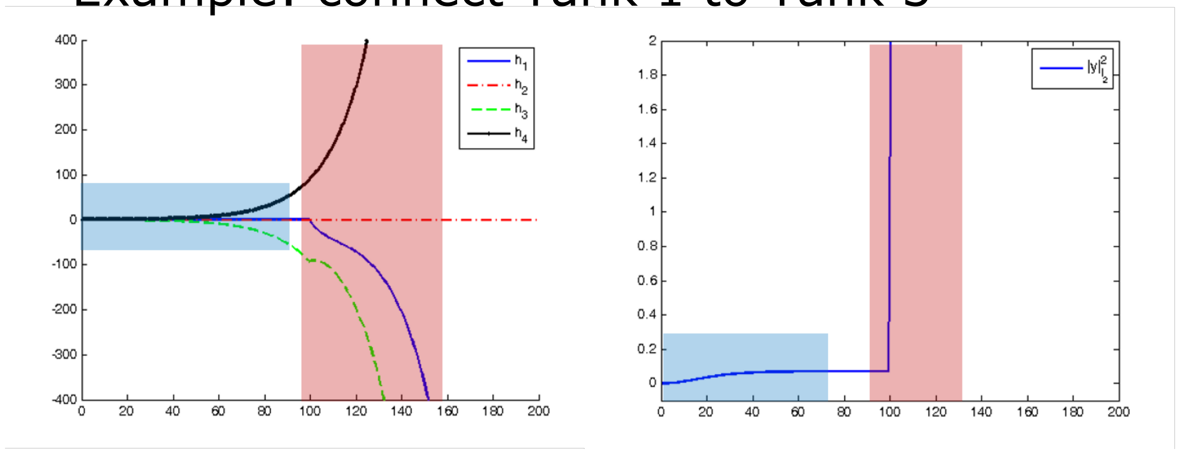
- Attack characterized by:

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

- Moving Target Defense (MTD):
    - Modify the system dynamics



Example: connect Tank 1 to Tank 3



Teixeira et al, "Revealing Stealthy Attacks in Control Systems". Allerton 2012

MTD creates uncertainty in the adversary.

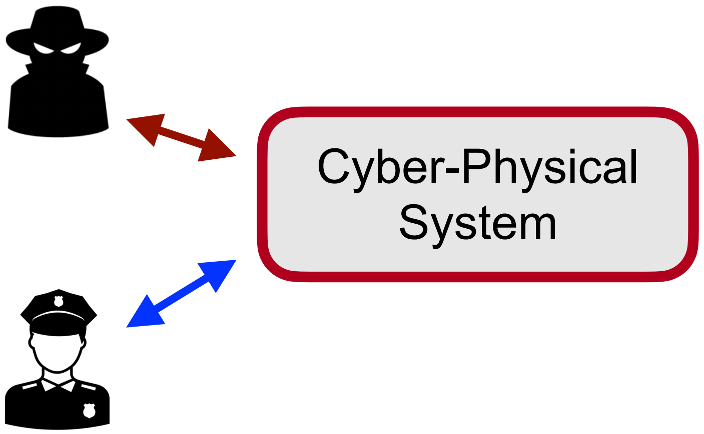
Interaction resembles a Stackelberg game
1. The attack policy is fixed according to the attacker's goals
2. MTD is implemented for detection (defender's goals)

Defense: Active Detection of Attacks

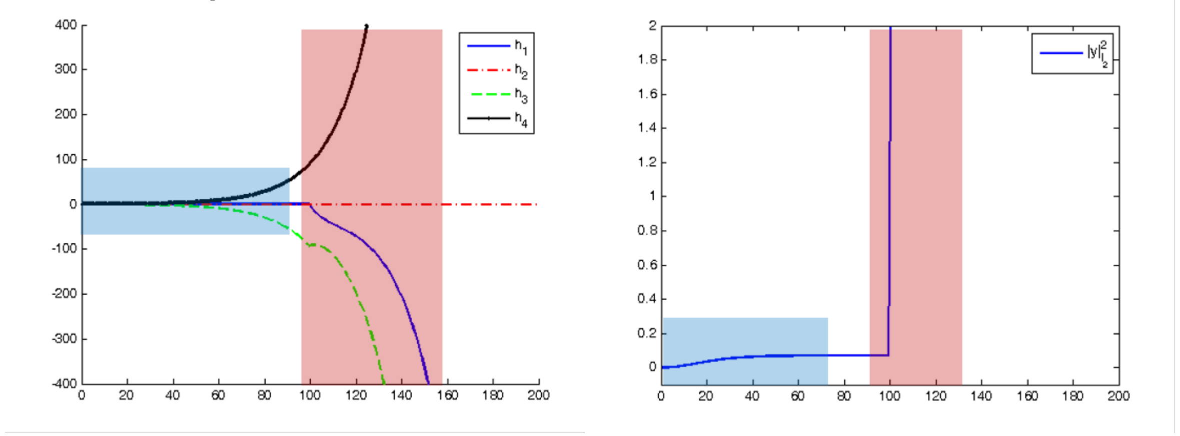
- Attack characterized by:

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

- Moving Target Defense (MTD):
    - Modify the system dynamics



Example: connect Tank 1 to Tank 3



Teixeira et al, “Revealing Stealthy Attacks in Control Systems”. Allerton 2012

MTD creates uncertainty in the adversary.

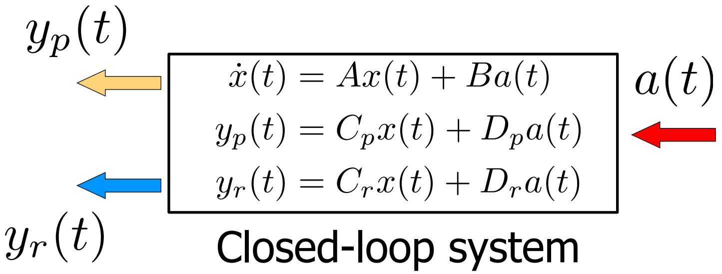
Interaction resembles a Stackelberg game
1. The attack policy is fixed according to the attacker’s goals
2. MTD is implemented for detection (defender’s goals)

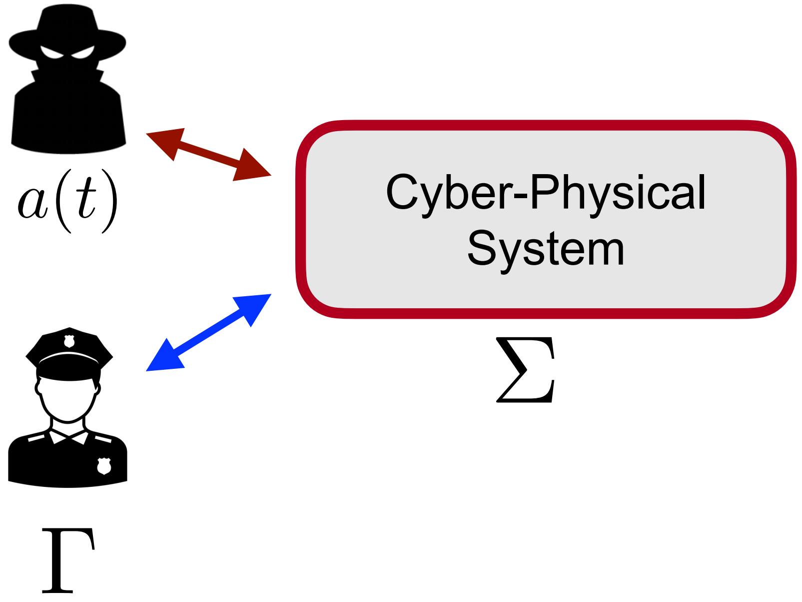
The setting is uninformative of how secure the ‘new’ system is against ‘new’ attack policies!

Outline

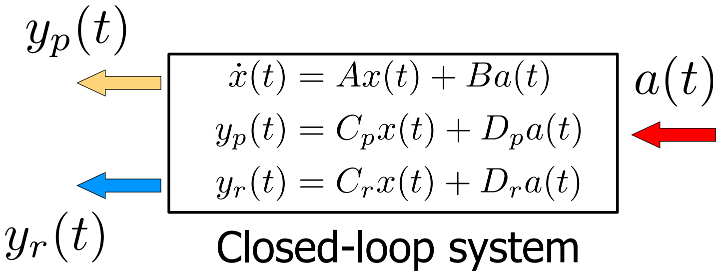
- Security Risk Management
- Scenario and Threat Models
- **Security Metrics and Game-Theoretic Design**
- Security under Model Uncertainty
- Probabilistic Risk Measures and Game-Theoretic Design
- Conclusions and Remarks

Security Metrics and Game-Theoretic Design

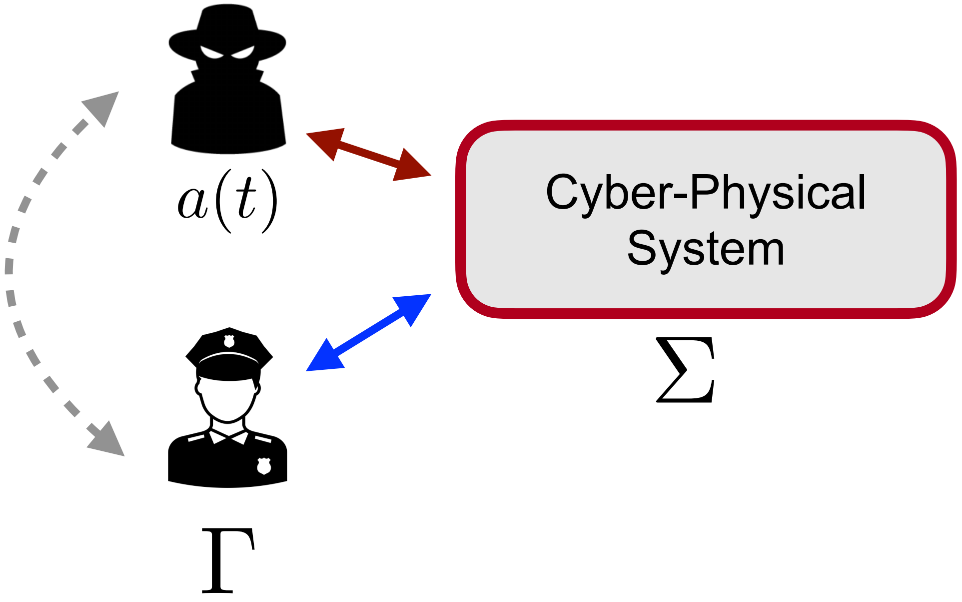




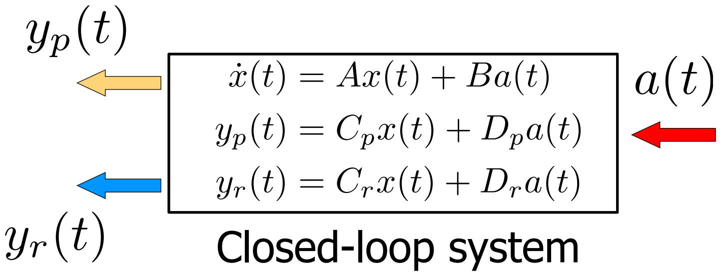
Security Metrics and Game-Theoretic Design



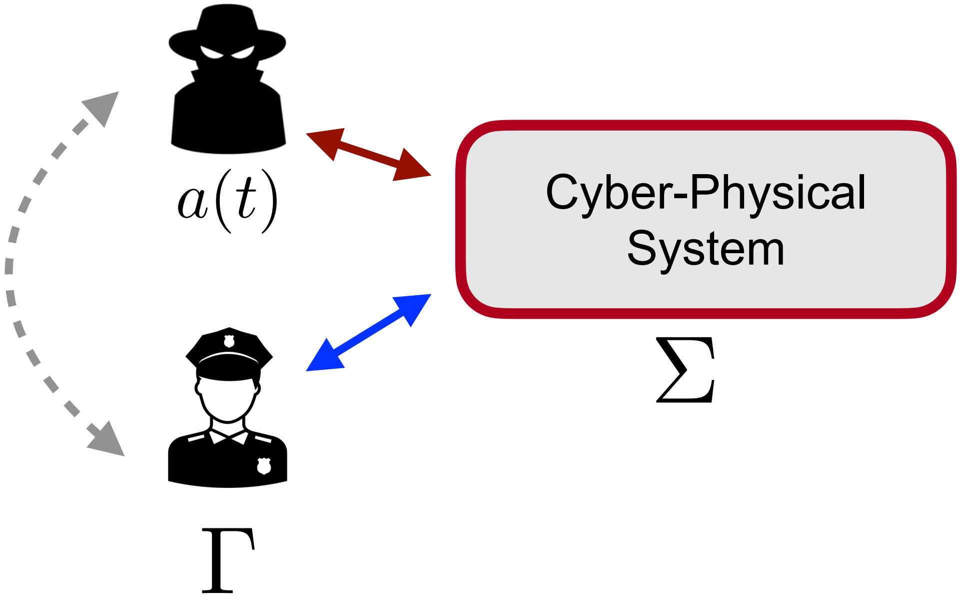
$$\min_{\Gamma} \max_{a} J(\Gamma, a; \Sigma)$$



Security Metrics and Game-Theoretic Design

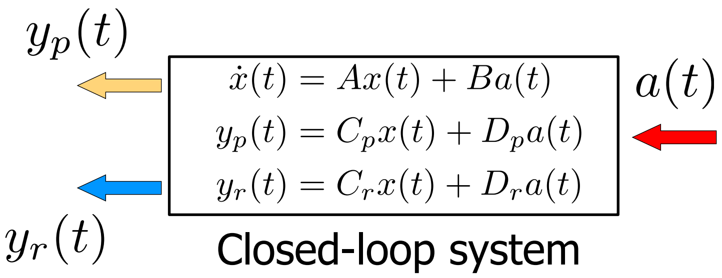


$$\min_{\Gamma} \max_{a} J(\Gamma, a; \Sigma)$$

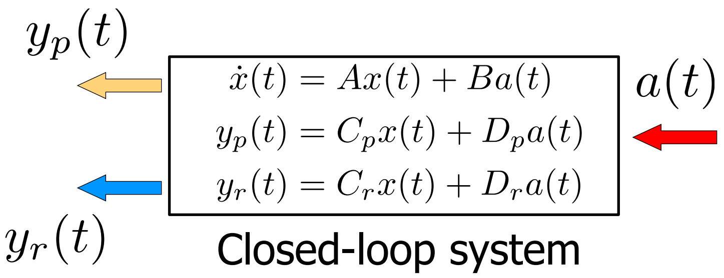


The security metric $J(\Gamma, a; \Sigma)$ captures the interactions between attacker and defender through the system. This enables us to construct a richer set of games between these players.

Security Metric for Control Systems



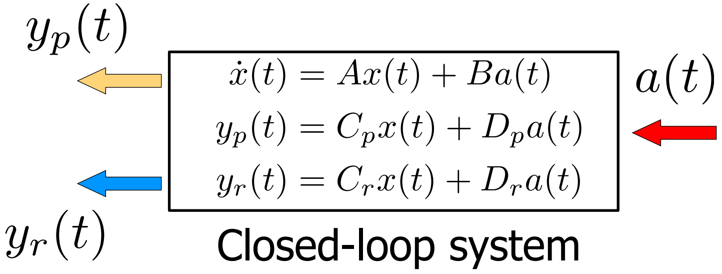
Security Metric for Control Systems



Attack policy: Maximise the impact on performance without raising alarms

Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

Security Metric for Control Systems



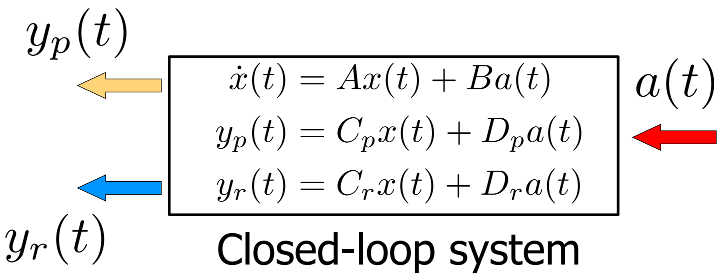
Attack policy: Maximise the impact on performance without raising alarms

Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

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

[Teixeira et al., CDC 15], [Teixeira, Springer 21]

Security Metric for Control Systems



Attack policy: Maximise the impact on performance without raising alarms

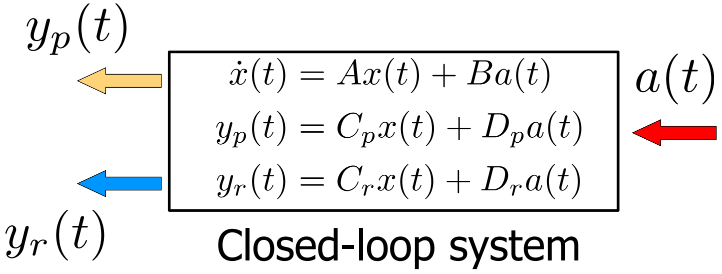
Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

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

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

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

[Teixeira et al., CDC 15], [Teixeira, Springer 21]



Security Metric for Control Systems

Attack policy: Maximise the impact on performance without raising alarms

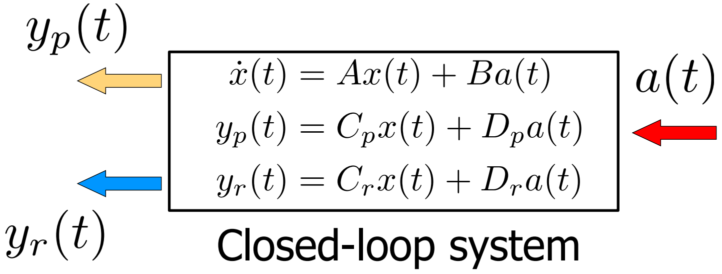
Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

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

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

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

- 'Unstable zero dynamics' is an optimal policy



Security Metric for Control Systems

Attack policy: Maximise the impact on performance without raising alarms

Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

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

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

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

- 'Unstable zero dynamics' is an optimal policy

- An equivalent formulation (dual problem):

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

[Teixeira et al., CDC 15], [Teixeira, Springer 21]

Security Metric for Control Systems

$y_p(t)$ ← Closed-loop system ← $a(t)$

$y_r(t)$ ←

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

Closed-loop system

Attack policy: Maximise the impact on performance without raising alarms

Output-to-output gain: Maximize $\|y_p\|$, while keeping $\|y_r\|$ small

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

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

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

- 'Unstable zero dynamics' is an optimal policy

- An equivalent formulation (dual problem):

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

- Finite $\gamma^\star$ implies a bound on the performance degradation by stealthy attacks

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

[Teixeira et al., CDC 15], [Teixeira, Springer 21]

Security Analysis through Linear Matrix Inequalities

Security Analysis through Linear Matrix Inequalities

An equivalent formulation:

$$\gamma^\star = \min_{\beta \geq 0} \quad \beta$$

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

(infinite-dimensional constraint)

Security Analysis through Linear Matrix Inequalities

An equivalent formulation:

$$\gamma^\star = \min_{\beta \geq 0} \quad \beta$$

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

(infinite-dimensional constraint)

The constraint can be re-cast as a Linear Matrix Inequality (LMI)

- Key technique: Dissipative Systems Theory (details in backup slides)

Can be efficiently solved by SDP solvers (e.g., through CVX)

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

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

[Teixeira et al., CDC 15], [Teixeira, Springer 21]

Game-Theoretic Design through Bilinear Matrix Inequalities

Design problem for a Controller (L) and a Detector (K):

- K and L change the matrices of the closed-loop system

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

Game-Theoretic Design through Bilinear Matrix Inequalities

Design problem for a Controller (L) and a Detector (K):

- K and L change the matrices of the closed-loop system

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

Designing a Controller (L) & Detector (K) is possible under specific forms

… but leads to Bilinear Matrix Inequalities [Teixeira, Springer 2021] [Anand and Teixeira, IFAC WC 2020]

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

Game-Theoretic Design through Bilinear Matrix Inequalities

Design problem for a Controller (L) and a Detector (K):

- K and L change the matrices of the closed-loop system

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

Designing a Controller (L) & Detector (K) is possible under specific forms

… but leads to Bilinear Matrix Inequalities [Teixeira, Springer 2021] [Anand and Teixeira, IFAC WC 2020]

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

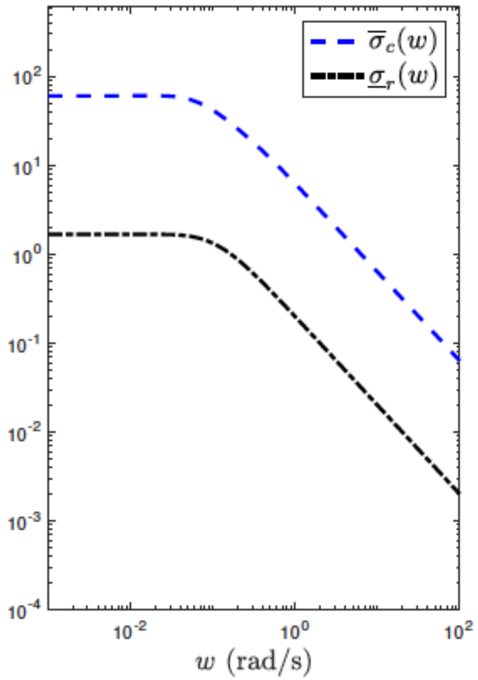
Next we use an heuristic: alternating minimisation

Example 1: Continuous-time

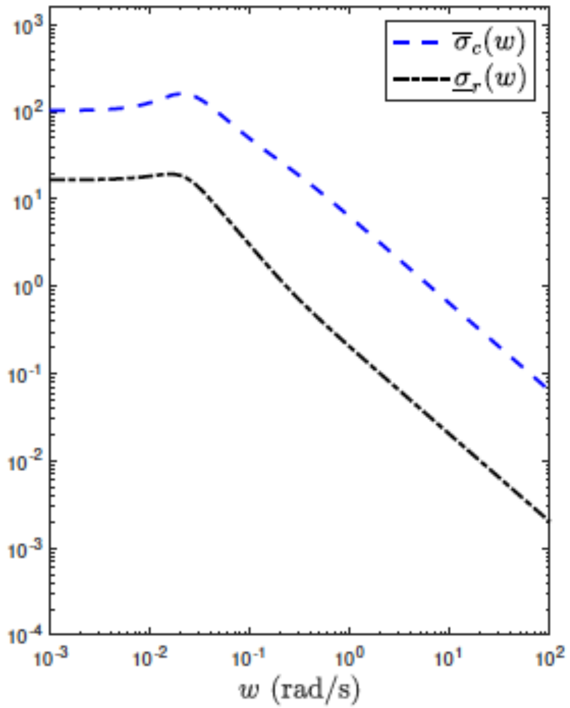
[Teixeira, Springer 2021]

- Classical vs Re-designed controller and detector

Nominal Design



After re-design



Example 2: Discrete-time

[Anand and Teixeira, IFAC WC 2020]

- Classical vs Re-designed controller and detector

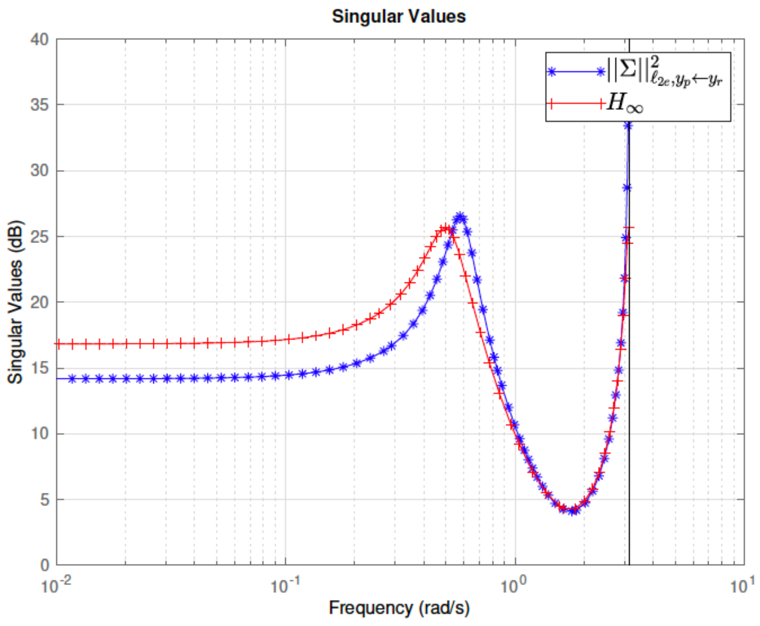


Fig. 1. Singular values - Performance output $(\bar{\sigma}(\Sigma_p))$

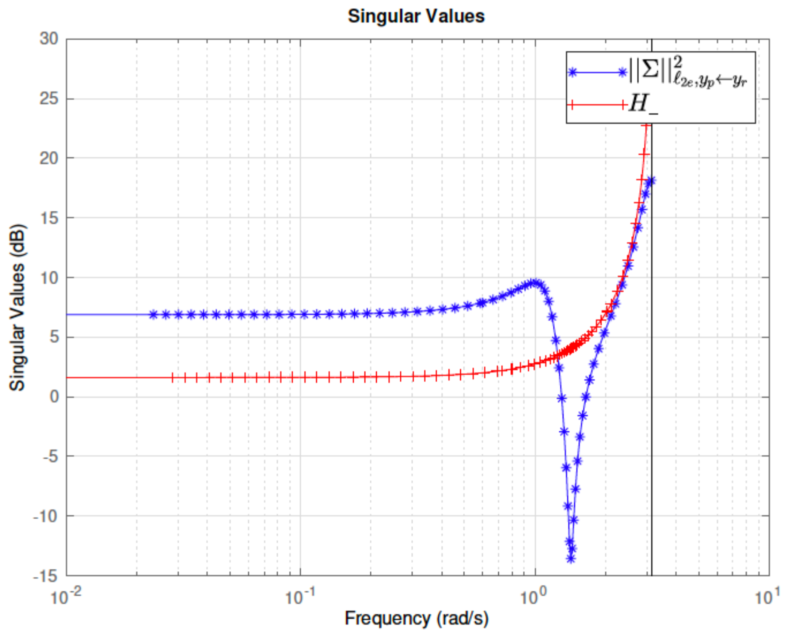
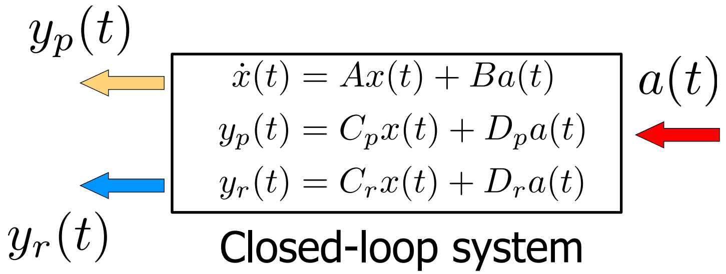


Fig. 2. Singular values - Detection output $(\underline{\sigma}(\Sigma_r))$

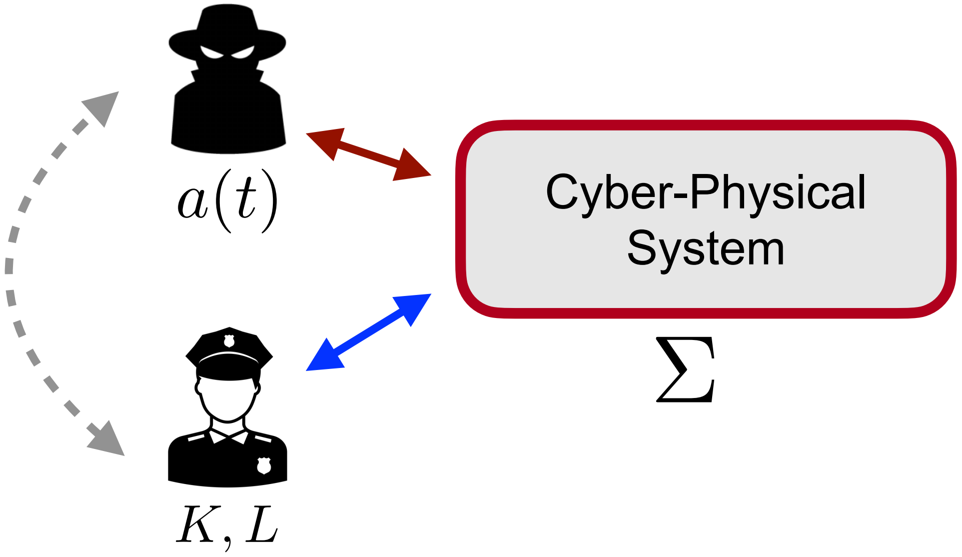
Security Metrics and Game-Theoretic Design



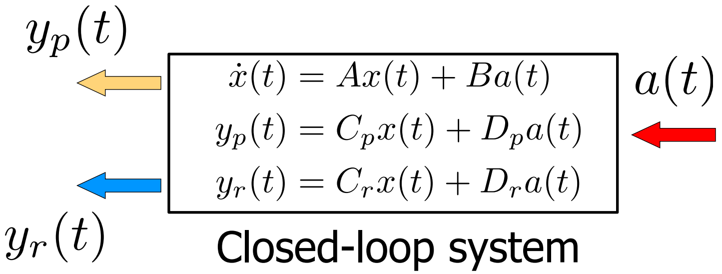
$$\min_{K,L} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_p\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_r\|^2_{\mathcal{L}_2} \leq 1$$

$$x(0) = 0$$



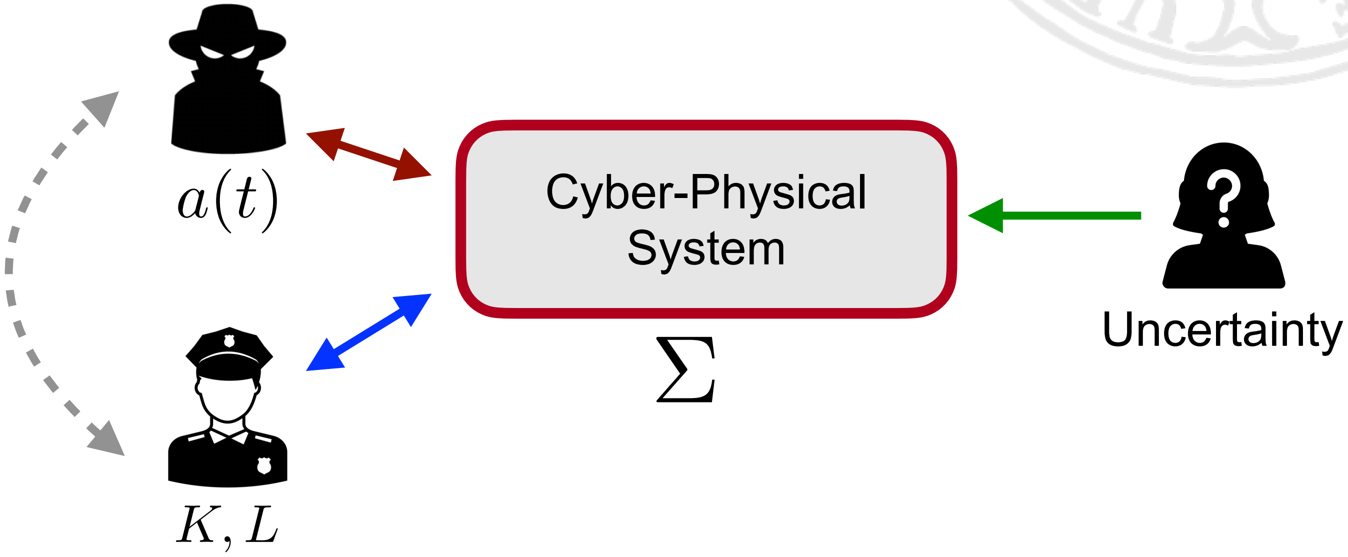
Security Metrics and Game-Theoretic Design



$$\min_{K,L} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_p\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_r\|^2_{\mathcal{L}_2} \leq 1$$

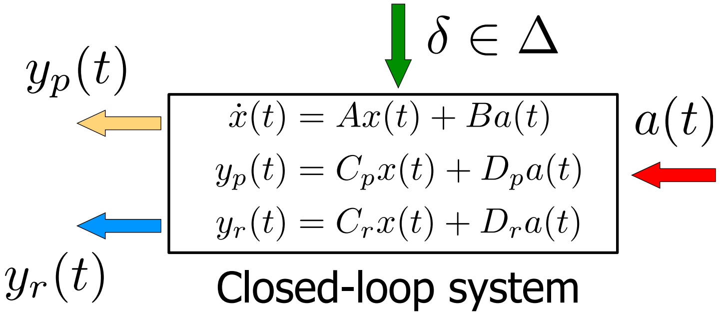
$$x(0) = 0$$

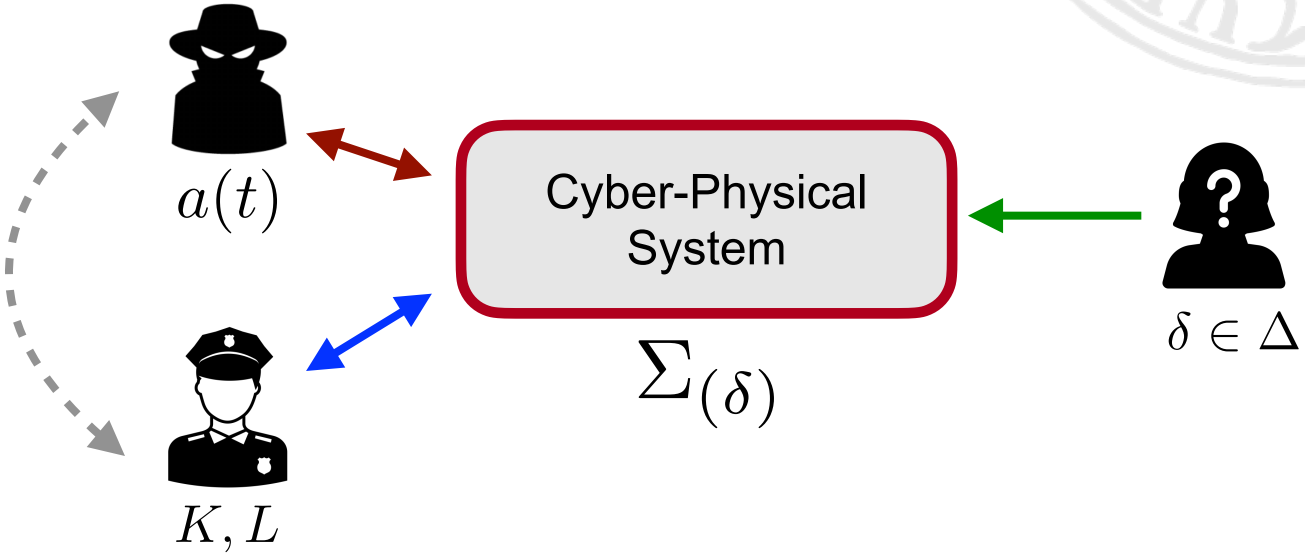


Outline

- Security Risk Management
- Scenario and Threat Models
- Security Metrics and Game-Theoretic Design
- **Security under Model Uncertainty**
- Probabilistic Risk Measures and Game-Theoretic Design
- Conclusions and Remarks

Security Metrics under Model Uncertainty

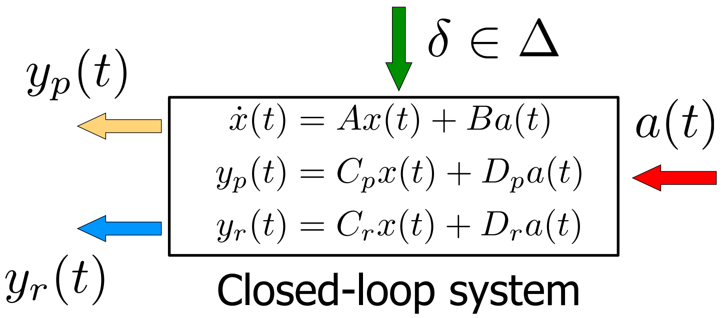


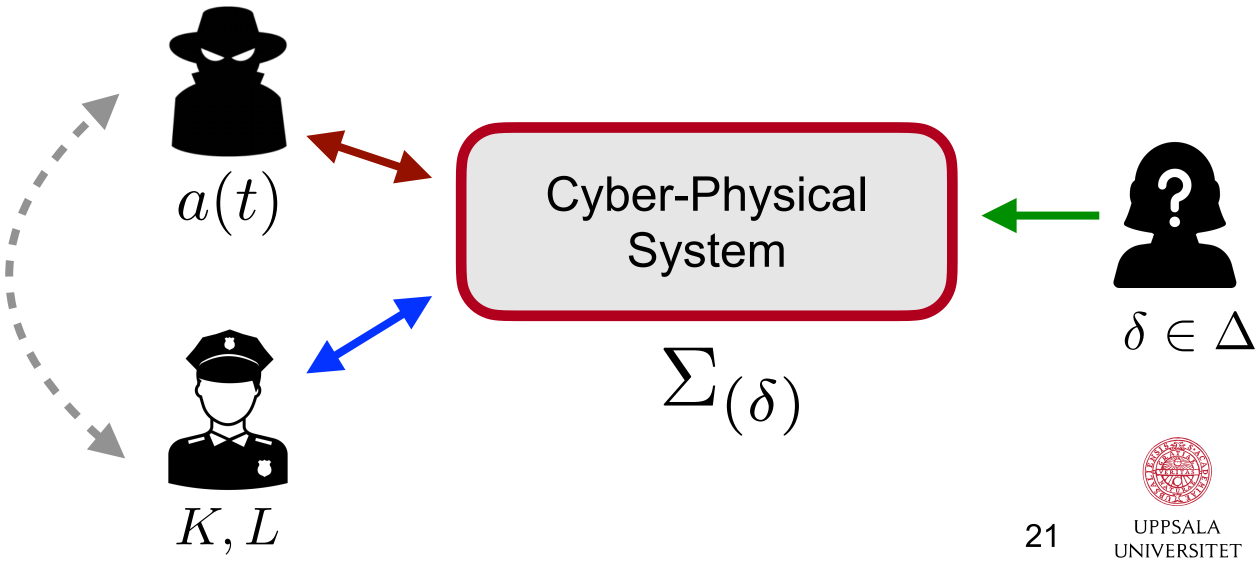


$$
\begin{aligned}
\sup_{a \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|^2_{\mathcal{L}_2} \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
$$

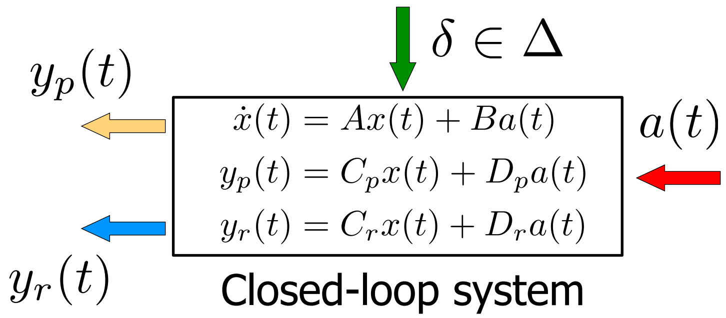
How should uncertainty be embedded in the Defender and Adversary?

Worst-Case Model Uncertainty





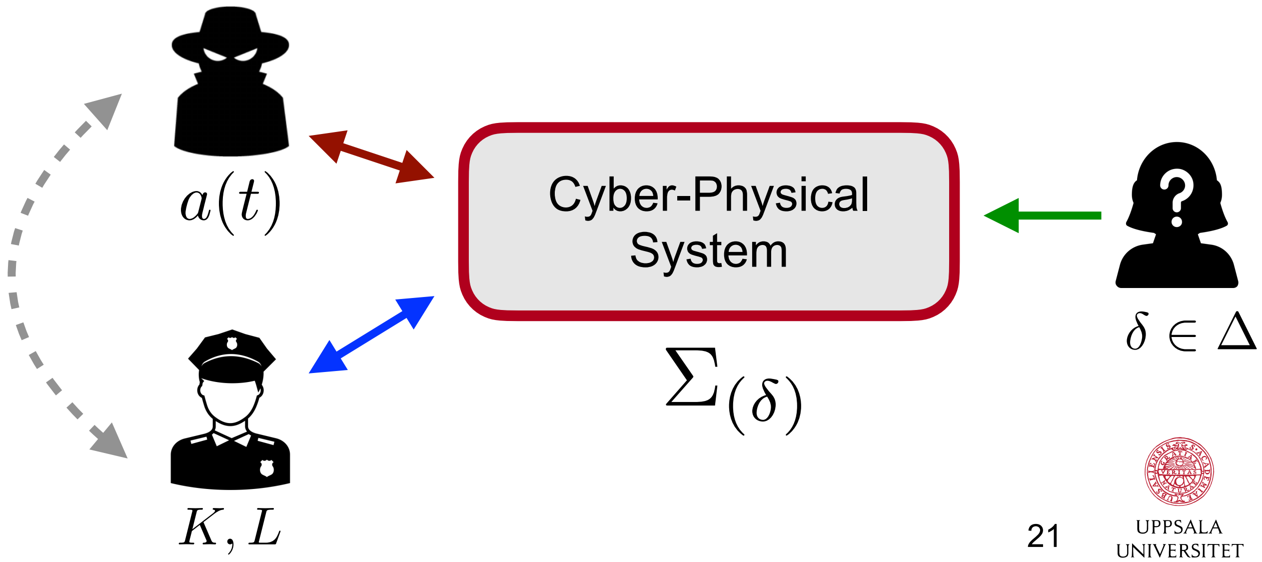
Worst-Case Model Uncertainty



$$\sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|_{\mathcal{L}_2}^2$$

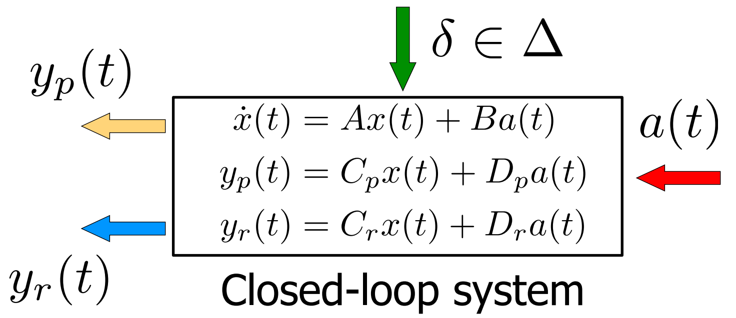
$$\text{s.t.} \quad \|y_{r,(\delta)}\|_{\mathcal{L}_2}^2 \leq 1$$

$$x_{(\delta)}(0) = 0$$



UPPSALA UNIVERSITET

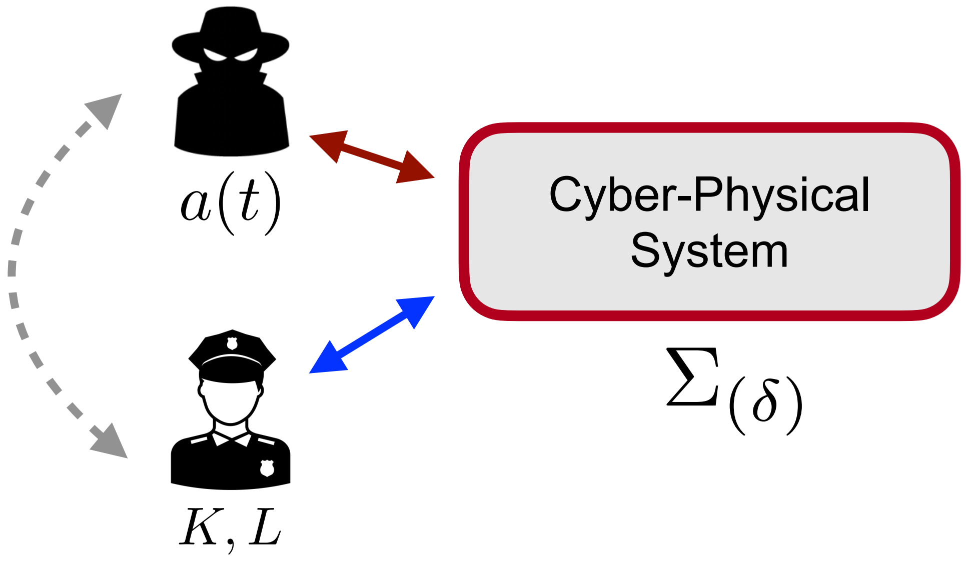
Worst-Case Model Uncertainty



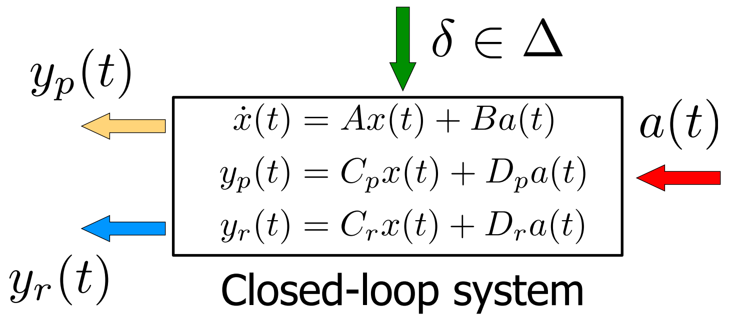
$$\sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

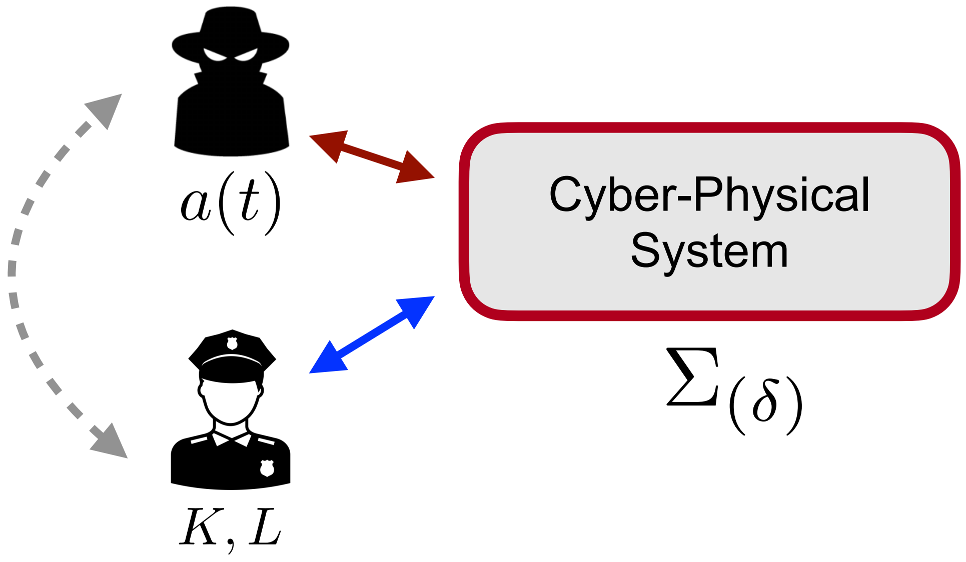
$\Leftrightarrow$

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



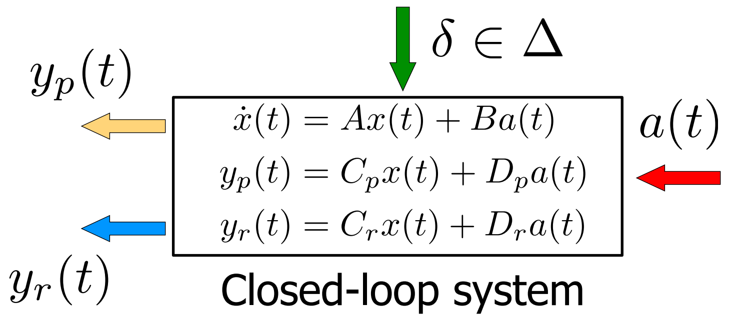
Worst-Case Model Uncertainty

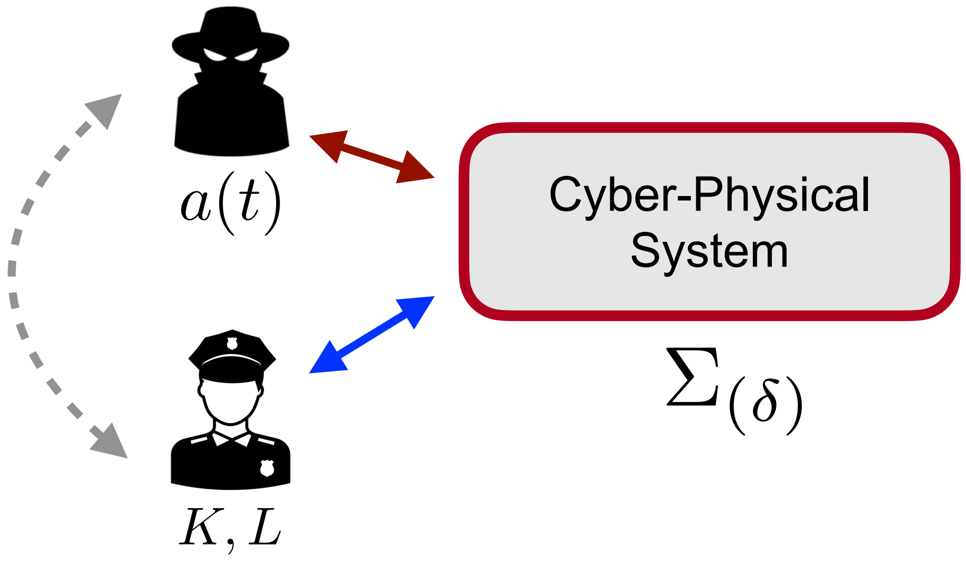




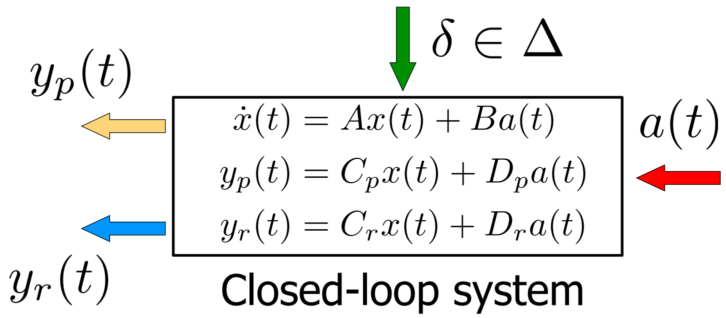
Worst-Case Model Uncertainty

$$\sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$





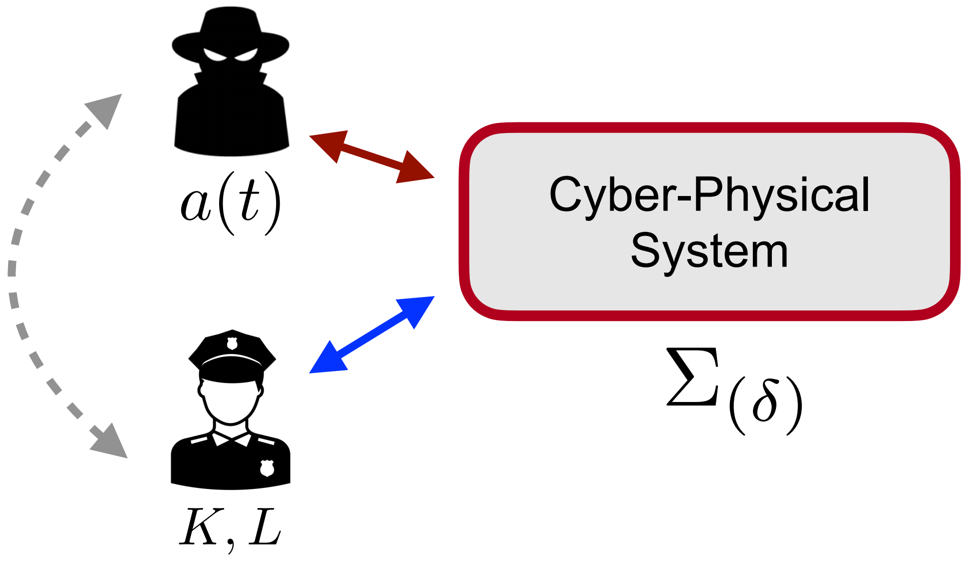
Worst-Case Model Uncertainty



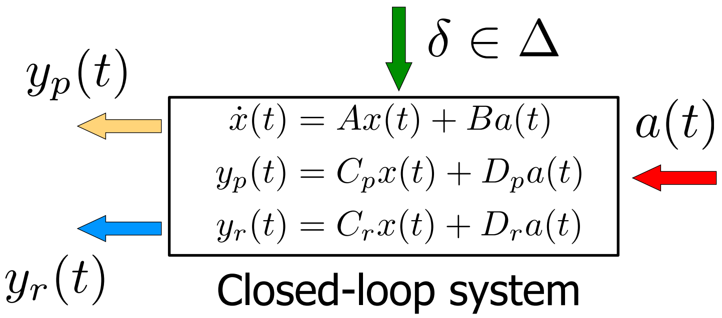
$$\sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

$\Longleftrightarrow$

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



Worst-Case Model Uncertainty



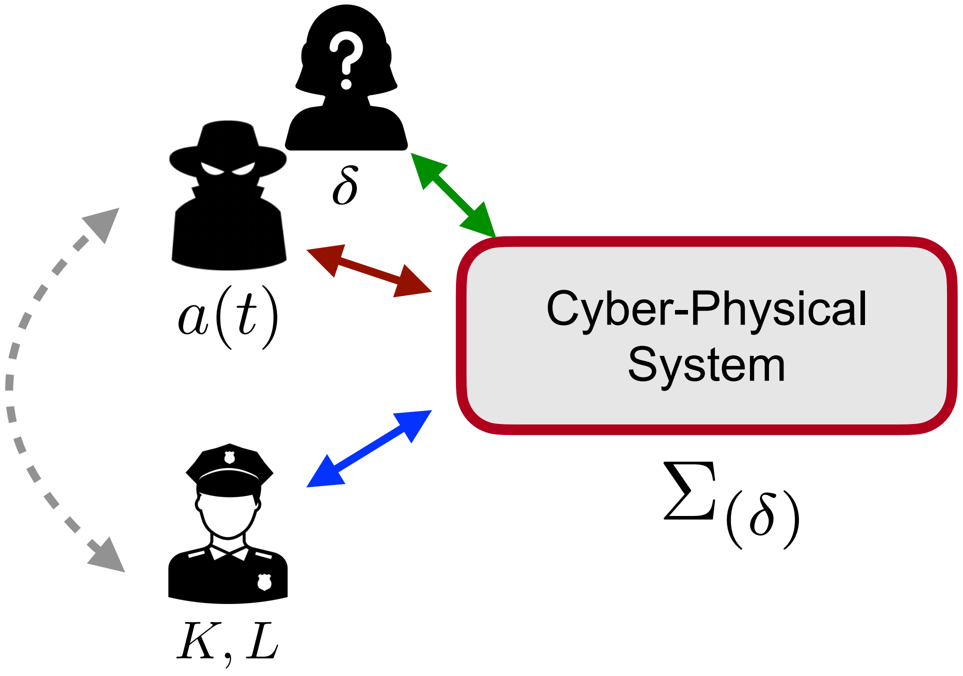
$$\begin{aligned} \sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|^2_{\mathcal{L}_2} \\ \text{s.t.} \quad & \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1 \\ & x_{(\delta)}(0) = 0 \end{aligned}$$

$\Longleftrightarrow$

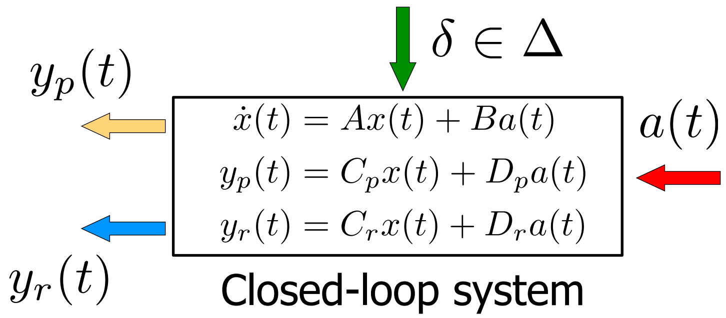
$$\begin{aligned} \gamma^\star = \min_{\beta \geq 0} \quad & \beta \\ \text{s.t.} \quad & \beta \|y_r\|^2_{\mathcal{L}_2} - \|y_p\|^2_{\mathcal{L}_2} \geq 0, \ \forall \, a \in \mathcal{L}_{2e}, \ \forall \delta \in \Delta \end{aligned}$$

Adversary and Uncertainty are colluding!

Uncertainty “reacts” to defender’s actions



Worst-Case Model Uncertainty



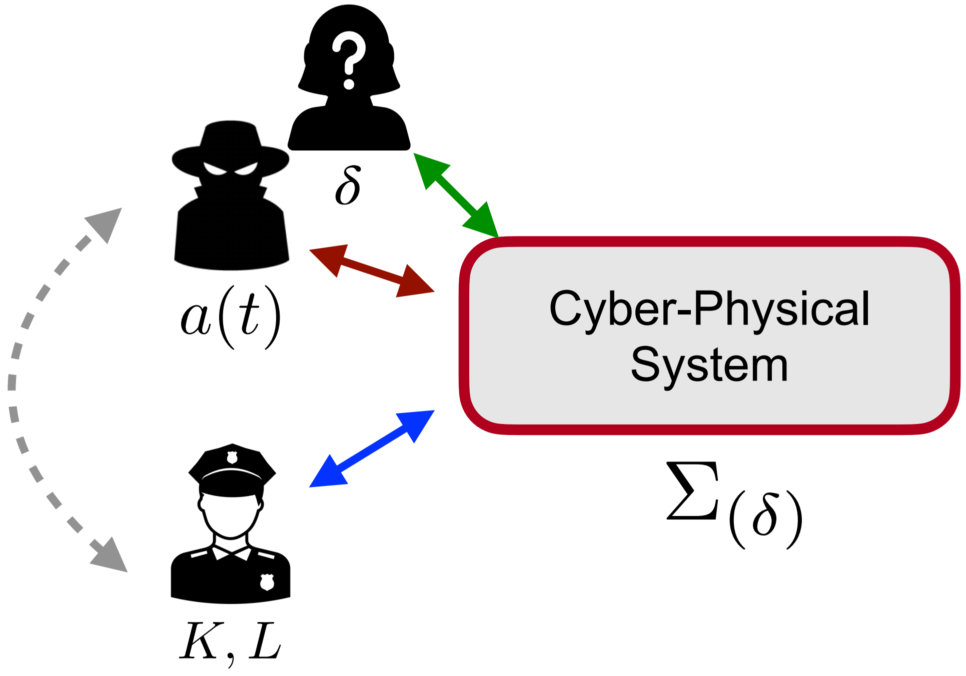
$$
\begin{aligned}
\sup_{\delta \in \Delta} \sup_{a \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|_{\mathcal{L}_2}^2 \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|_{\mathcal{L}_2}^2 \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
\quad \Longleftrightarrow \quad
\begin{aligned}
\gamma^\star = \min_{\beta \geq 0} \quad & \beta \\
\text{s.t.} \quad & \beta \|y_r\|_{\mathcal{L}_2}^2 - \|y_p\|_{\mathcal{L}_2}^2 \geq 0, \ \forall\, a \in \mathcal{L}_{2e}, \ \forall \delta \in \Delta
\end{aligned}
$$

Adversary and Uncertainty are colluding!

Uncertainty “reacts” to defender’s actions

As in robust control, worst-case disturbance can be conservative!

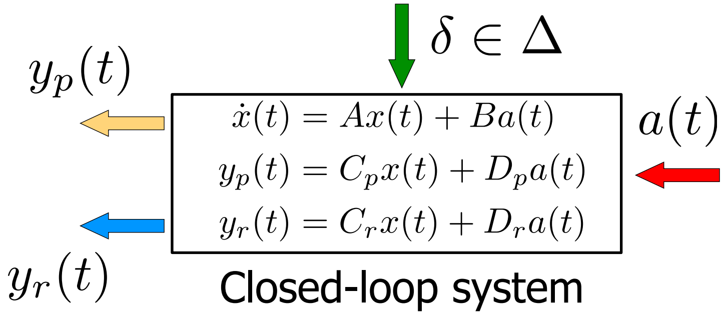
G. C. Calafiore and M. C. Campi, "The scenario approach to robust control design," *IEEE TAC*, 2006

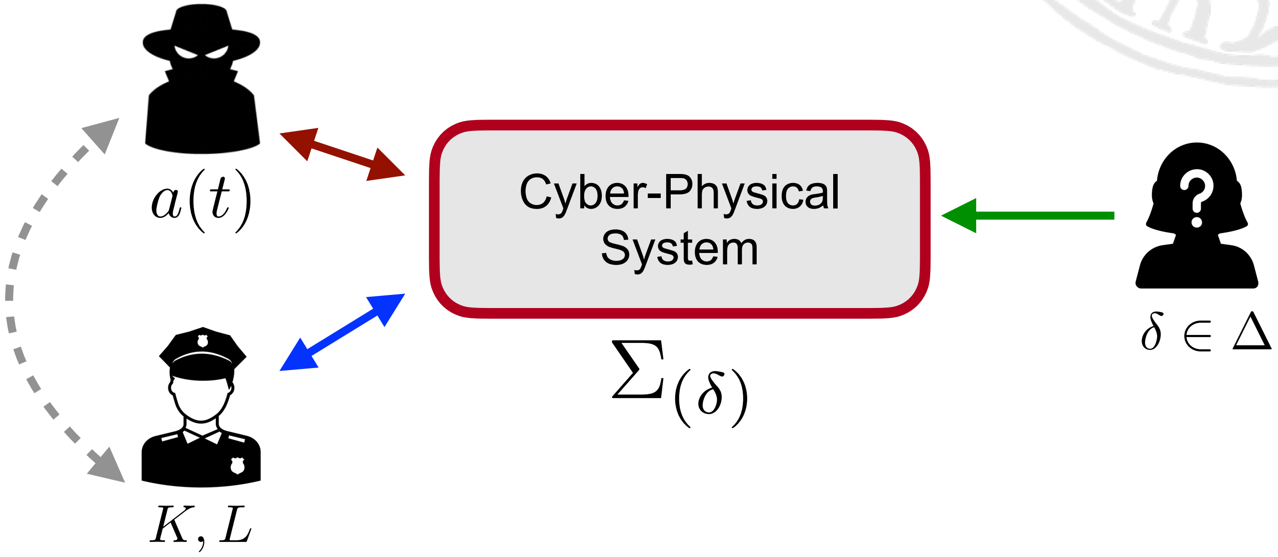


Outline

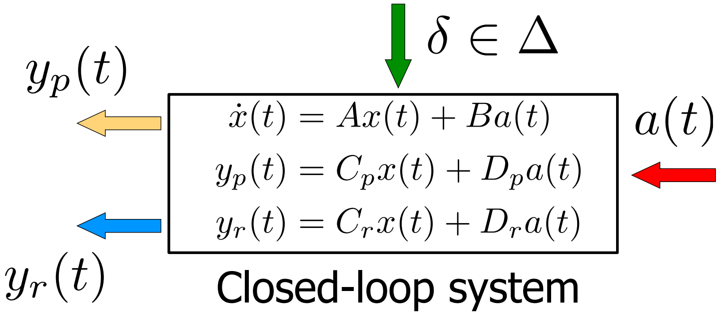
- Security Risk Management
- Scenario and Threat Models
- Security Metrics and Game-Theoretic Design
- Security under Model Uncertainty
- **Probabilistic Risk Measures and Game-Theoretic Design**
- Conclusions and Remarks

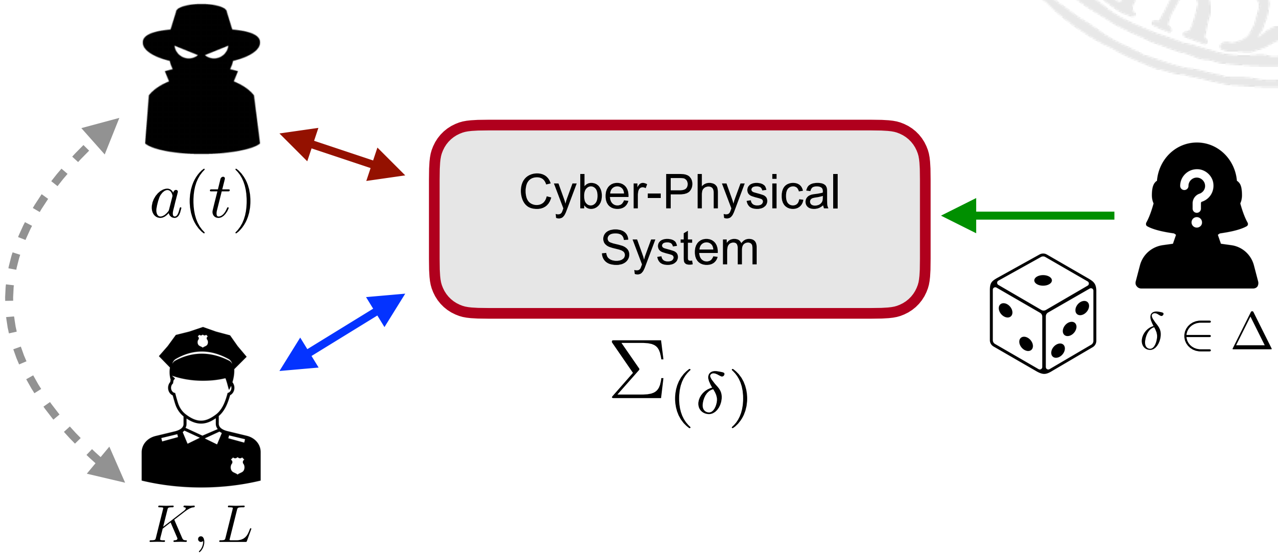
Security Metrics under Probabilistic Model Uncertainty



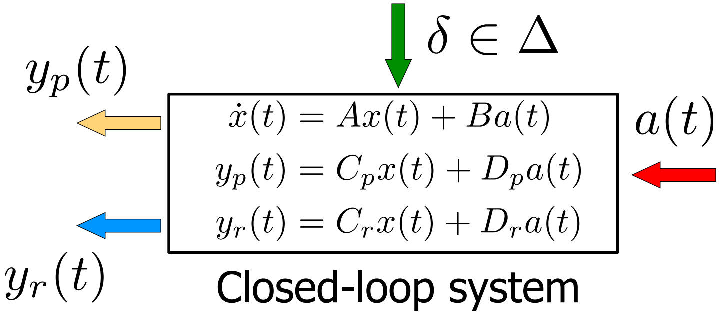


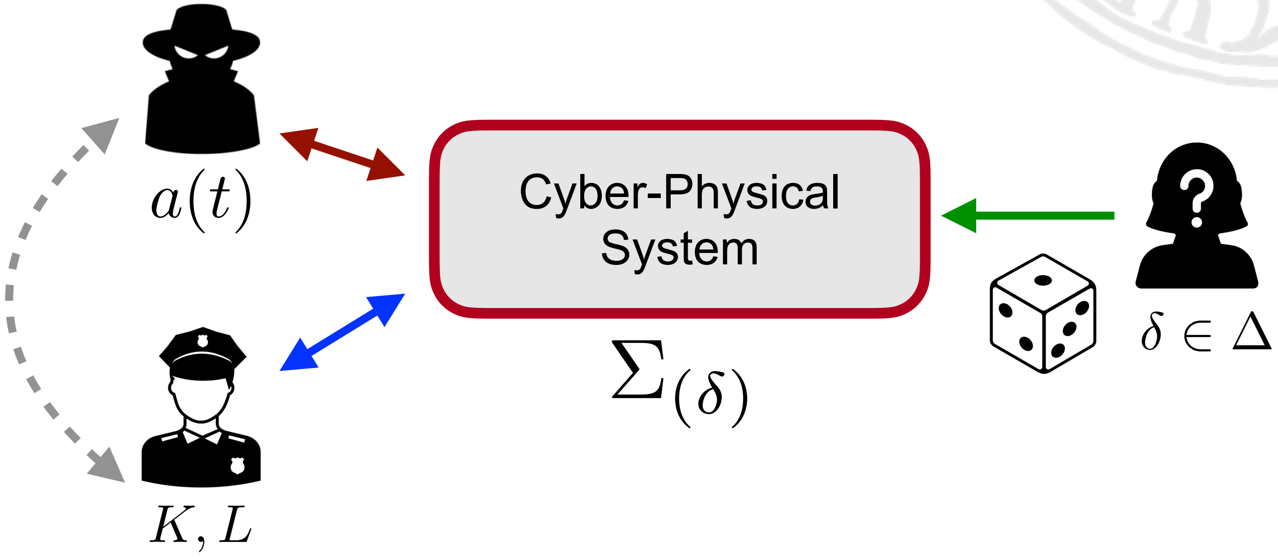
Security Metrics under Probabilistic Model Uncertainty





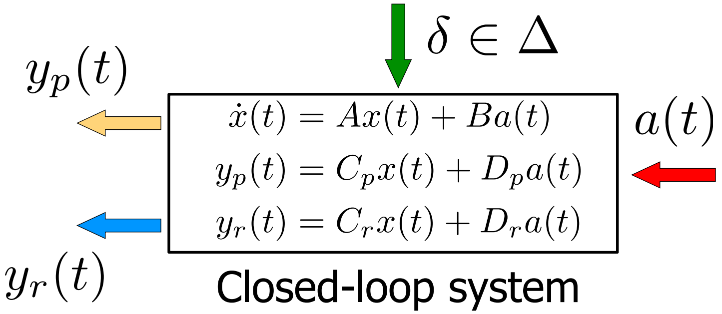
Security Metrics under Probabilistic Model Uncertainty

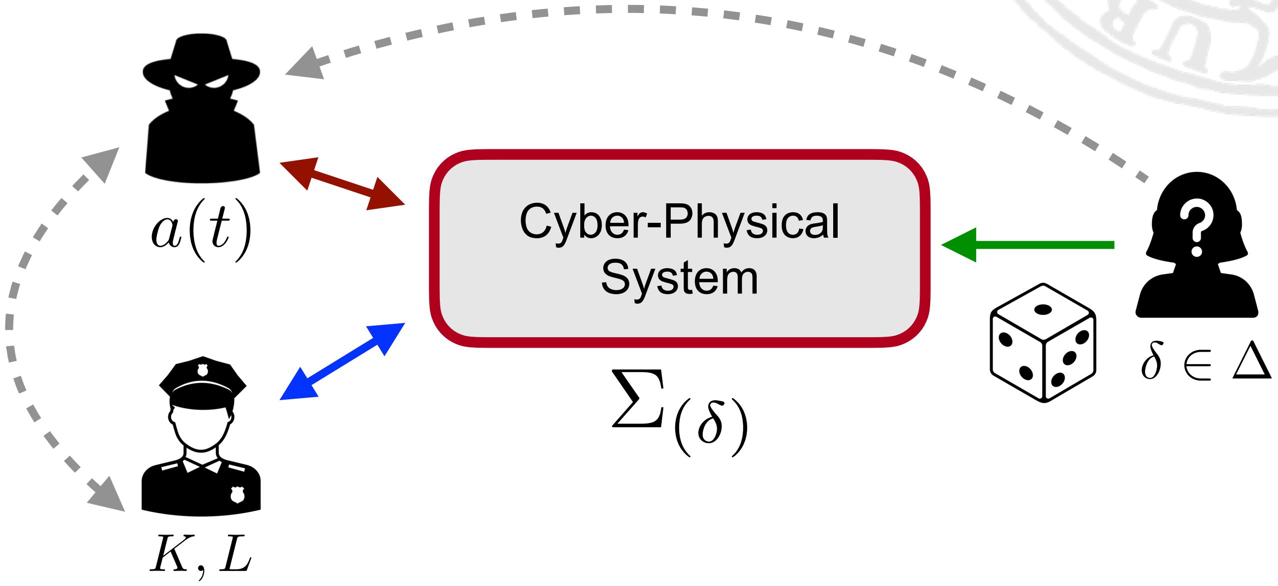




What is the information structure between the Uncertainty and the Adversary?

Security Metrics under Probabilistic Model Uncertainty



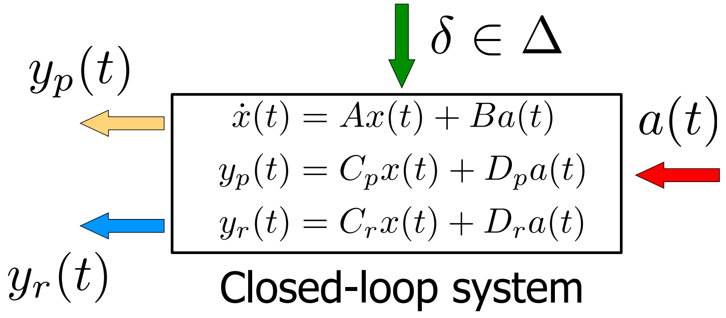


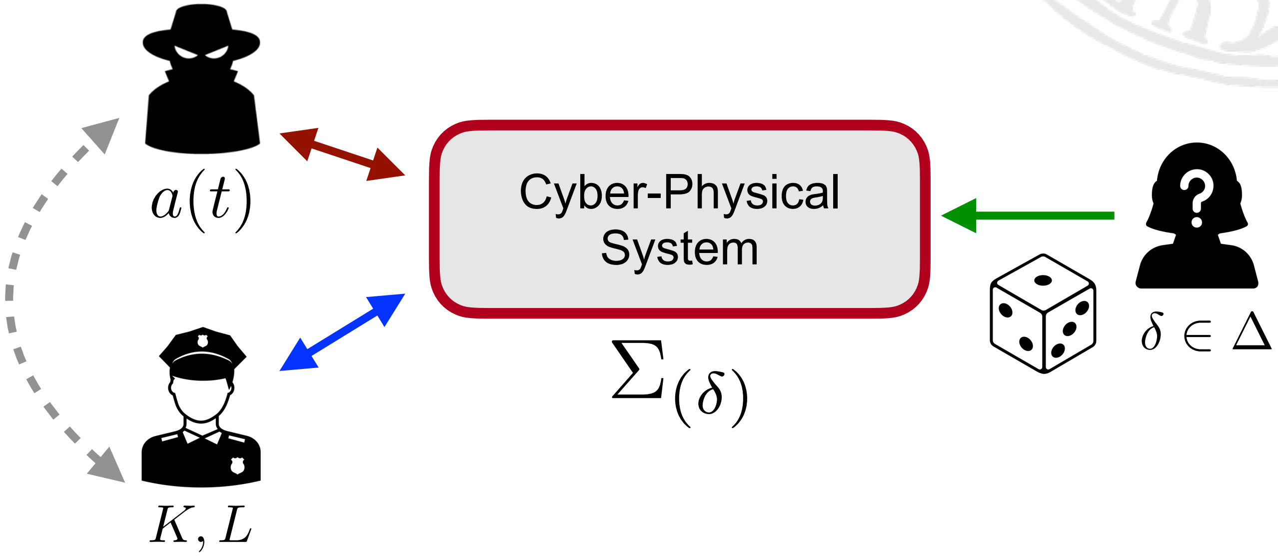
What is the information structure between the Uncertainty and the Adversary?

(i) **Omniscient Adversary:** knows the realization of the uncertainty, but they do not collude.

(i) Anand, Teixeira. "Risk-based Security Measure Allocation Against Actuator Attacks". IEEE Open Journal of Control Systems, 2023.

Security Metrics under Probabilistic Model Uncertainty





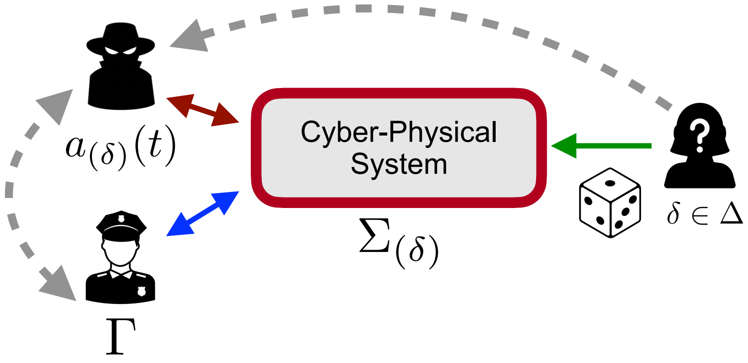
What is the <u>information structure</u> between the Uncertainty and the Adversary?

(i) **Omniscient Adversary:** knows the realization of the uncertainty, but they do not collude.

(ii) **Imperfect-information Adversary:** does not know the realization, needs to be robust to the uncertainty

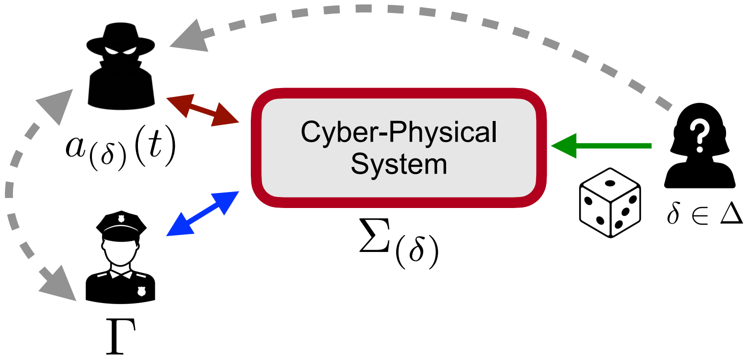
(i) Anand, Teixeira. "Risk-based Security Measure Allocation Against Actuator Attacks". IEEE Open Journal of Control Systems, 2023.
(ii) Anand et al.. "Risk Assessment of Stealthy Attacks on Uncertain Control Systems". IEEE TAC, 2023

Probabilistic Risk Measures and Game-Theoretic Design



UPPSALA UNIVERSITET

Probabilistic Risk Measures and Game-Theoretic Design

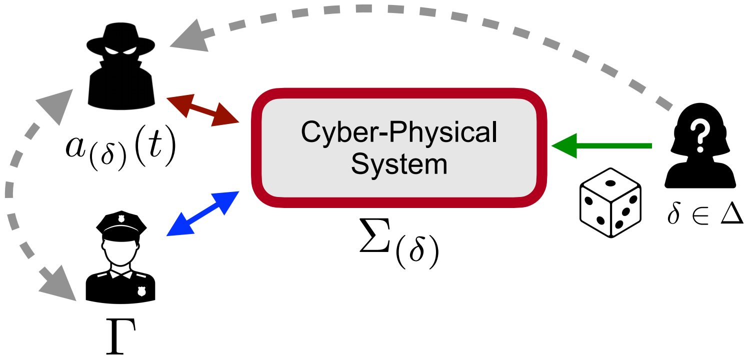


Impact of an **Omniscient Adversary:**

$$
\begin{aligned}
q(\Gamma, \delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|^2_{\mathcal{L}_2} \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
$$

Anand, Teixeira. “Risk-based Security Measure Allocation Against Actuator Attacks”. IEEE Open Journal of Control Systems, 2023.

Probabilistic Risk Measures and Game-Theoretic Design



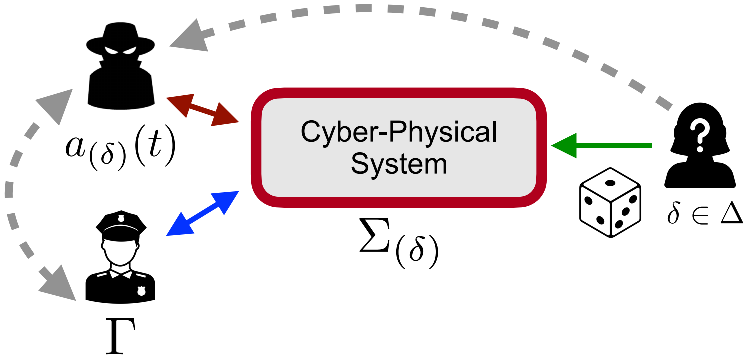
Impact of an **Omniscient Adversary:**

$$q(\Gamma,\delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

The impact $q(\Gamma,\delta)$ is a random variable with a distribution induced by the uncertainty.

Anand, Teixeira. “Risk-based Security Measure Allocation Against Actuator Attacks”. IEEE Open Journal of Control Systems, 2023.

Probabilistic Risk Measures and Game-Theoretic Design

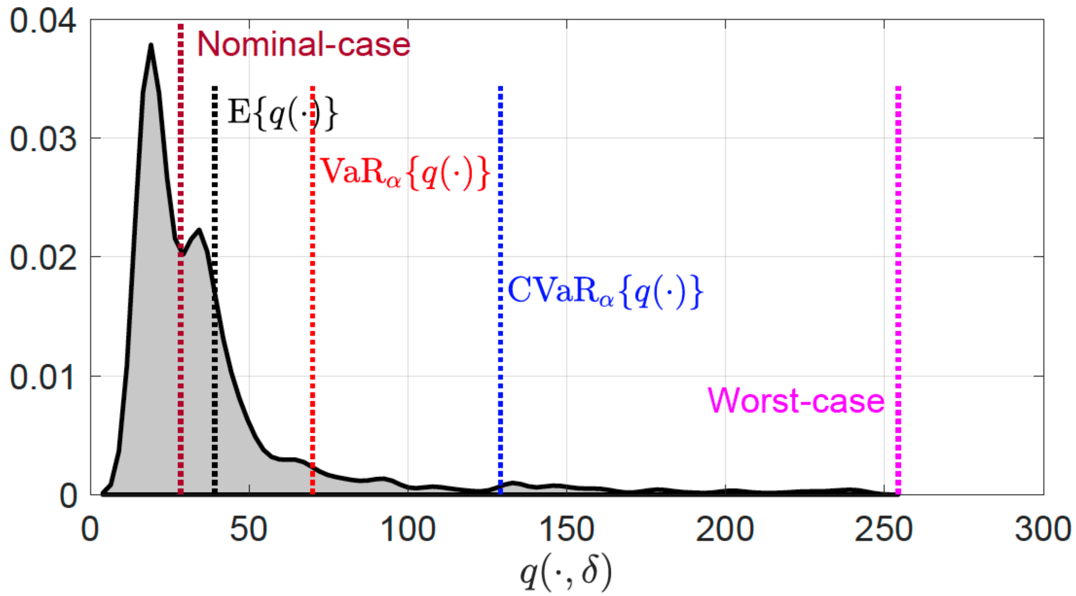


Impact of an **Omniscient Adversary:**

$$q(\Gamma,\delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

The impact $q(\Gamma,\delta)$ is a random variable with a distribution induced by the uncertainty.

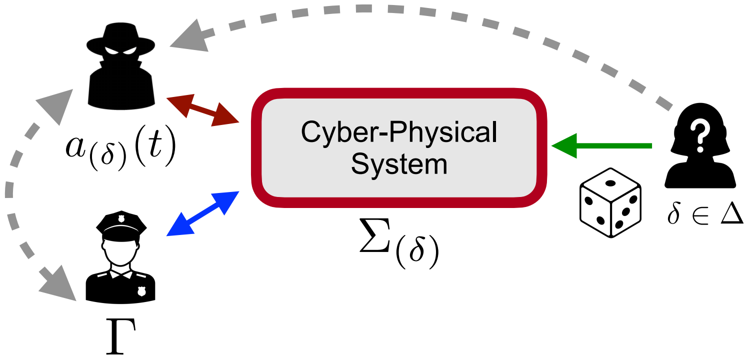
Need to define a **risk measure** $R_\Delta\{\,\cdot\,\}$ to marginalize away the uncertainty.



UPPSALA UNIVERSITET

Anand, Teixeira. "Risk-based Security Measure Allocation Against Actuator Attacks". IEEE Open Journal of Control Systems, 2023.

Probabilistic Risk Measures and Game-Theoretic Design



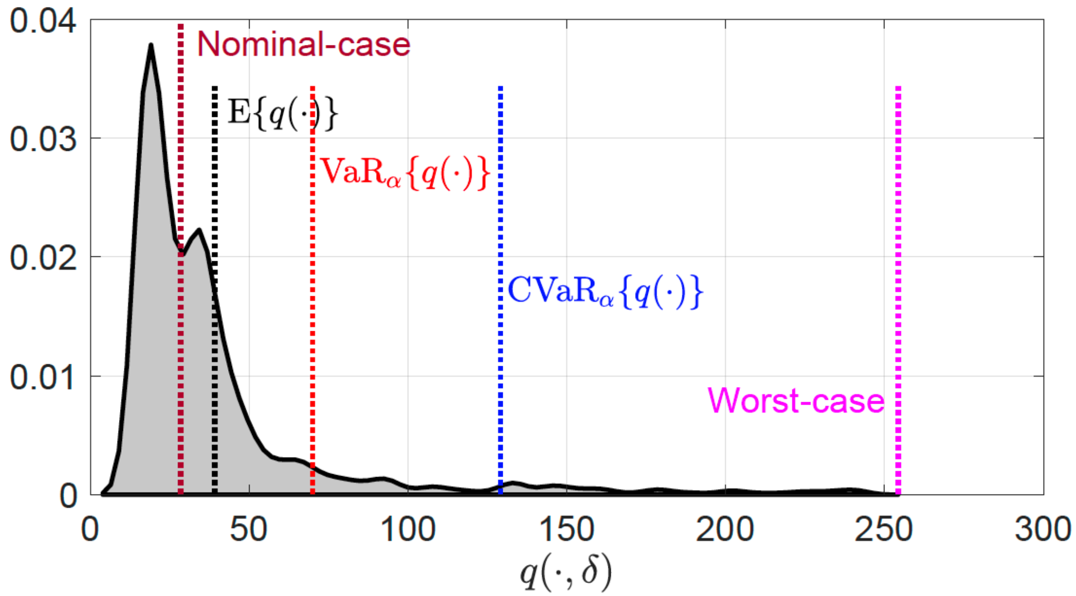
Impact of an **Omniscient Adversary:**

$$q(\Gamma, \delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

The impact $q(\Gamma, \delta)$ is a random variable with a distribution induced by the uncertainty.

Need to define a **risk measure** $R_\Delta\{\,\cdot\,\}$ to marginalize away the uncertainty.

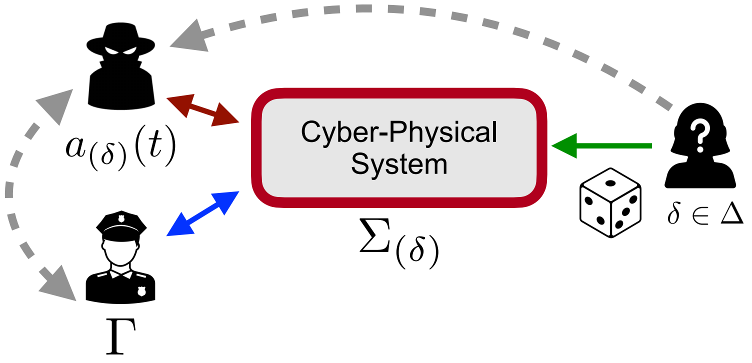
Use sample-based approximations.



UPPSALA UNIVERSITET

Anand, Teixeira. “Risk-based Security Measure Allocation Against Actuator Attacks”. IEEE Open Journal of Control Systems, 2023.

Probabilistic Risk Measures and Game-Theoretic Design



Impact of an **Omniscient Adversary:**

$$q(\Gamma,\delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

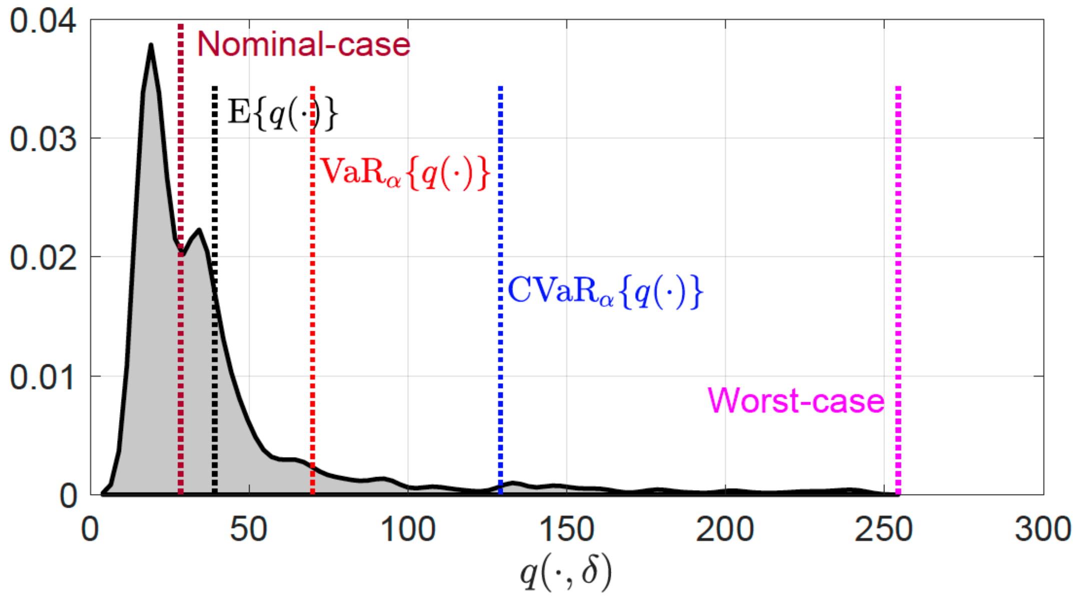
The impact $q(\Gamma,\delta)$ is a random variable with a distribution induced by the uncertainty.

Need to define a **risk measure** $R_\Delta\{\,\cdot\,\}$ to marginalize away the uncertainty.

Use sample-based approximations.

Risk-optimal defense:

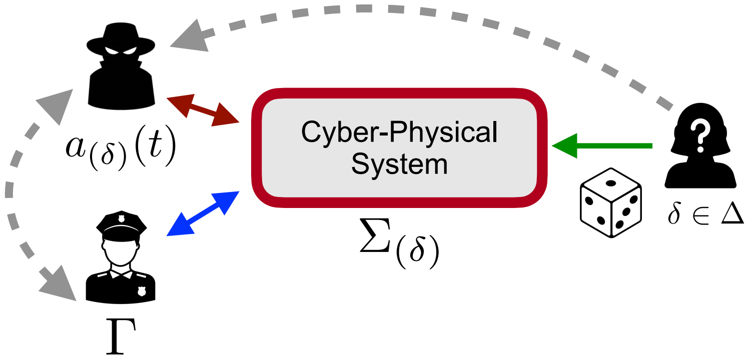
$$\min_{\Gamma} R_\Delta \{q(\Gamma,\delta)\}$$



UPPSALA UNIVERSITET

Anand, Teixeira. “Risk-based Security Measure Allocation Against Actuator Attacks”. IEEE Open Journal of Control Systems, 2023.

Example: allocation of protection on actuator channels

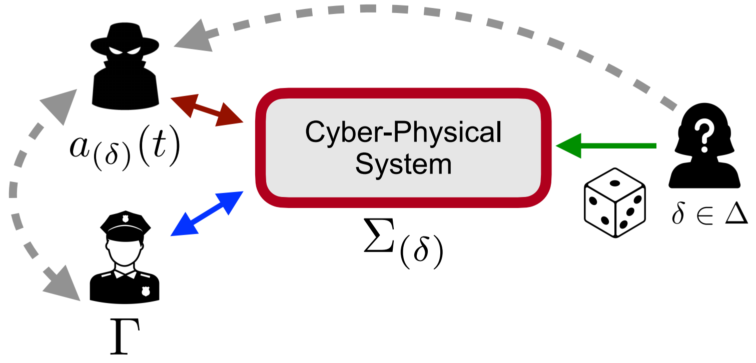


Γ - set of protected actuators

$$
\begin{aligned}
q(\Gamma, \delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|_{\mathcal{L}_2}^2 \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|_{\mathcal{L}_2}^2 \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
$$

UPPSALA UNIVERSITET

Example: allocation of protection on actuator channels



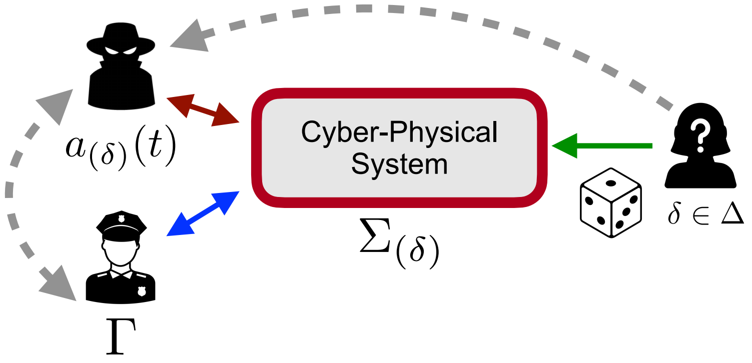
Γ - set of protected actuators

$$
\begin{aligned}
q(\Gamma, \delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|_{\mathcal{L}_2}^2 \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|_{\mathcal{L}_2}^2 \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
$$

The **risk measure** $R_{\Delta}\{\,\cdot\,\}$ is chosen as the CVaR, using sample-based approximations.

UPPSALA UNIVERSITET

Example: allocation of protection on actuator channels



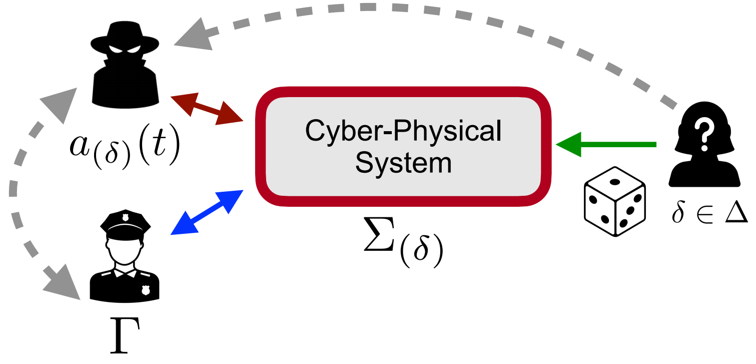
Γ - set of protected actuators

$$
\begin{aligned}
q(\Gamma, \delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad & \|y_{p,(\delta)}\|^2_{\mathcal{L}_2} \\
\text{s.t.} \quad & \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1 \\
& x_{(\delta)}(0) = 0
\end{aligned}
$$

The **risk measure** $R_\Delta\{\,\cdot\,\}$ is chosen as the CVaR, using sample-based approximations.

Risk-optimal defense: $\min_{|\Gamma| < n_w} R_\Delta \{q(\Gamma, \delta)\}$

Example: allocation of protection on actuator channels



Γ - set of protected actuators

$$q(\Gamma,\delta) \triangleq \sup_{a_{(\delta)} \in \mathcal{L}_{2e}} \quad \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$
$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1$$
$$x_{(\delta)}(0) = 0$$

The **risk measure** $R_\Delta\{\cdot\}$ is chosen as the CVaR, using sample-based approximations.

Risk-optimal defense: $\min_{|\Gamma|<n_w} R_\Delta\{q(\Gamma,\delta)\}$

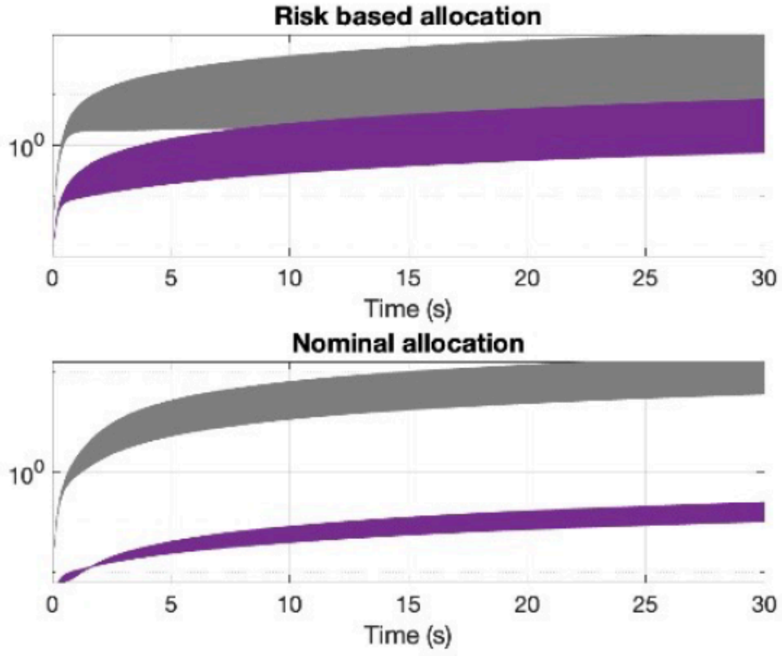


FIGURE 5. Performance energy (grey) and detection energy (violet) for $N = 500$ different realizations of uncertainty, under CVaR-based allocation strategy (top), and the nominal allocation strategy (bottom).

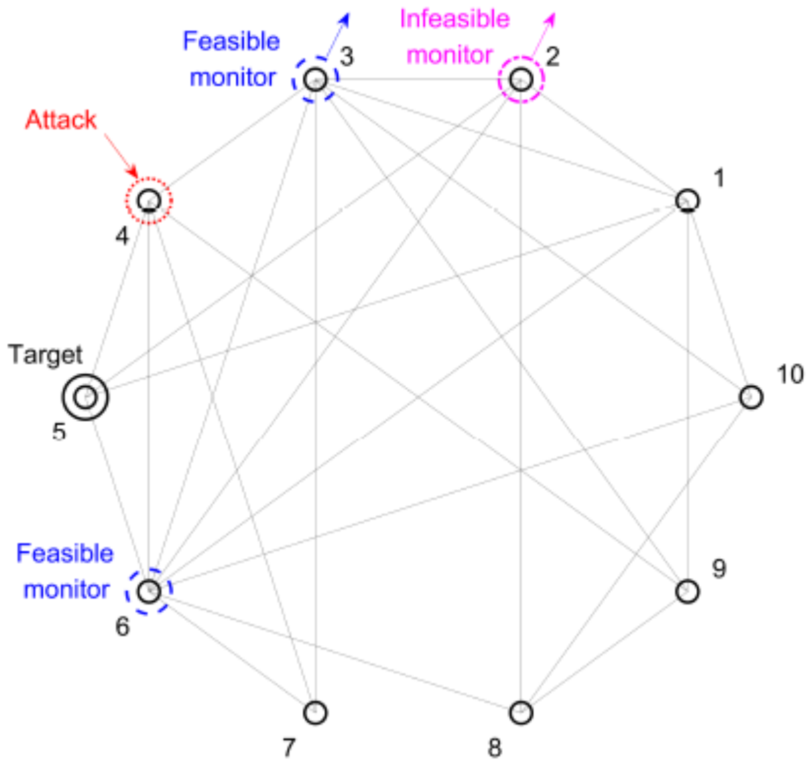
UPPSALA UNIVERSITET

Anand, Teixeira. “Risk-based Security Measure Allocation Against Actuator Attacks”. IEEE Open Journal of Control Systems, 2023.

Other Security Games with Omniscient Adversaries

Security Allocation in Uncertain Large-Scale Systems

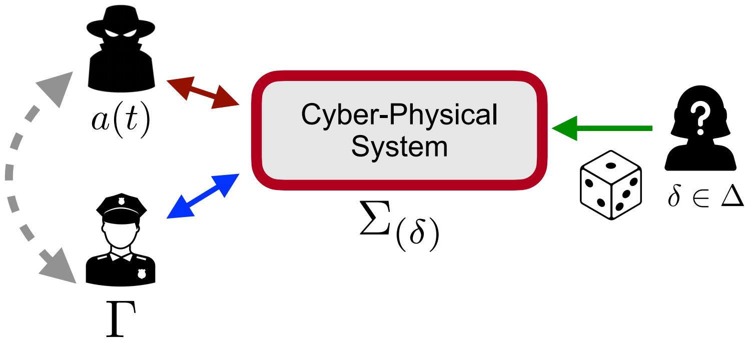
- Stackelberg games
  - Nguyen, et al. “Security Allocation in Networked Control Systems under Stealthy Attacks”. Submitted IEEE TCNS, 2023.
- Mixed Nash solutions and probabilistic uncertainty
  - Nguyen, et al. “A Zero-Sum Game Framework for Optimal Sensor Placement in Uncertain Networked Control Systems under Cyber-Attacks”. CDC 2022



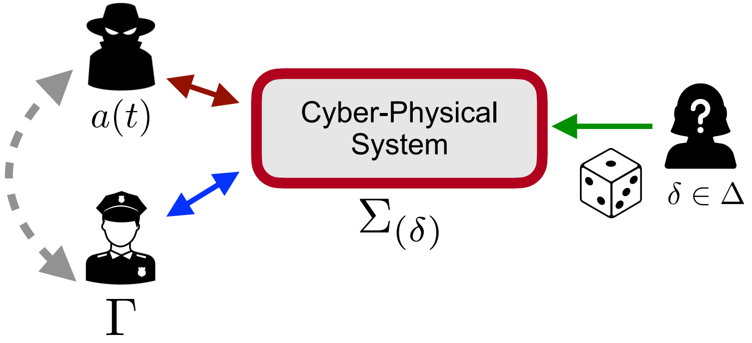
Risk-averse controller design

- Anand et al. “Risk-averse controller design against data injection attacks on actuators for uncertain control systems”. ACC 2022

Robust attacks under probabilistic uncertainty



Robust attacks under probabilistic uncertainty



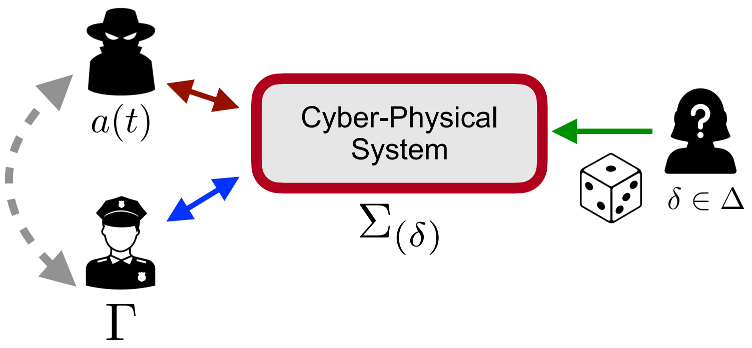
Risk of a **Imperfect-information Adversary:**

$$R_{\hat{\Delta}_N}(\Gamma) \triangleq \sup_{a \in \mathcal{L}_{2e}} \quad \frac{1}{N} \sum_{\delta \in \hat{\Delta}_N} \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1, \ \forall \delta \in \hat{\Delta}_N$$

Anand et al.. "Risk Assessment of Stealthy Attacks on Uncertain Control Systems". IEEE TAC, 2023

Robust attacks under probabilistic uncertainty



Risk of a **Imperfect-information Adversary:**

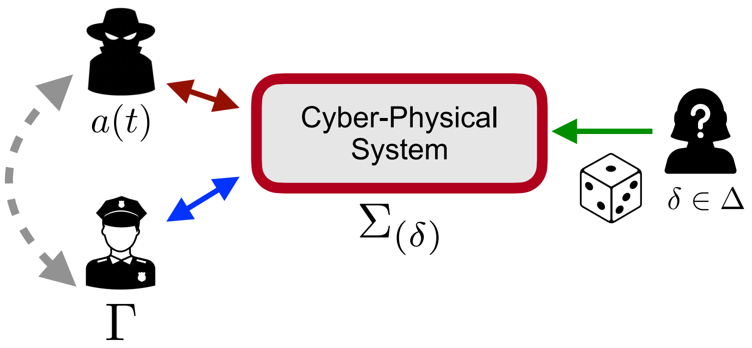
$$R_{\hat{\Delta}_N}(\Gamma) \triangleq \sup_{a \in \mathcal{L}_{2e}} \quad \frac{1}{N} \sum_{\delta \in \hat{\Delta}_N} \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1, \ \forall \delta \in \hat{\Delta}_N$$

The **risk measure** $R_{\hat{\Delta}_N}\{\,\cdot\,\}$ is computed based on N samples.

Anand et al.. "Risk Assessment of Stealthy Attacks on Uncertain Control Systems". IEEE TAC, 2023

Robust attacks under probabilistic uncertainty



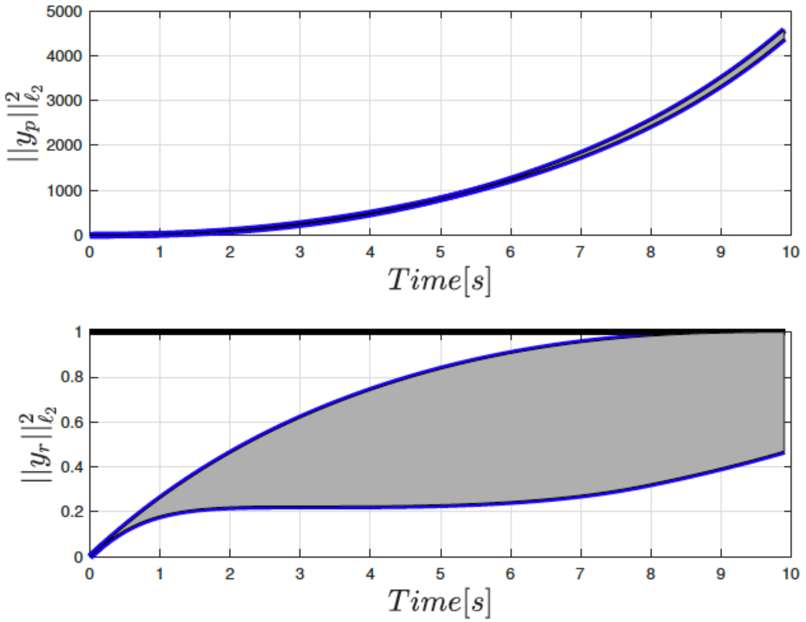
Risk of a **Imperfect-information Adversary:**

$$R_{\hat{\Delta}_N}(\Gamma) \triangleq \sup_{a \in \mathcal{L}_{2e}} \frac{1}{N} \sum_{\delta \in \hat{\Delta}_N} \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1, \ \forall \delta \in \hat{\Delta}_N$$

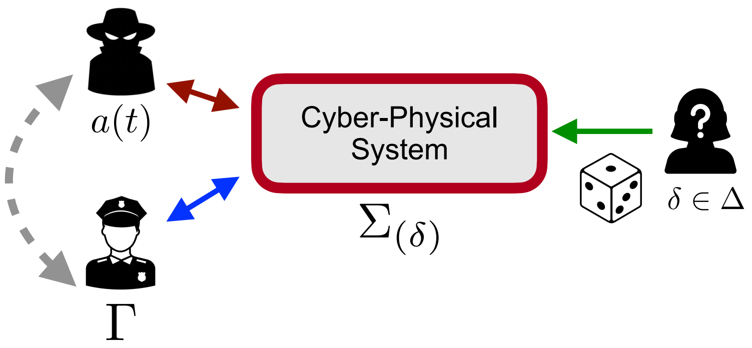
The **risk measure** $R_{\hat{\Delta}_N}\{\,\cdot\,\}$ is computed based on N samples.

The attack signal must be stealthy for all sampled realizations. Very likely to be bounded!



Anand et al.. “Risk Assessment of Stealthy Attacks on Uncertain Control Systems”. IEEE TAC, 2023

Robust attacks under probabilistic uncertainty



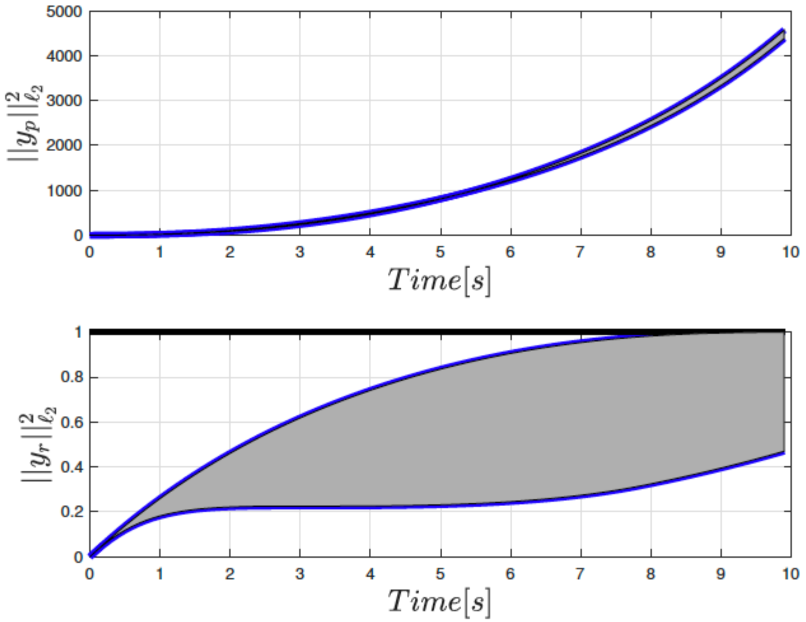
Risk of a **Imperfect-information Adversary:**

$$R_{\hat{\Delta}_N}(\Gamma) \triangleq \sup_{a \in \mathcal{L}_{2e}} \frac{1}{N} \sum_{\delta \in \hat{\Delta}_N} \|y_{p,(\delta)}\|^2_{\mathcal{L}_2}$$

$$\text{s.t.} \quad \|y_{r,(\delta)}\|^2_{\mathcal{L}_2} \leq 1, \; \forall \delta \in \hat{\Delta}_N$$

The **risk measure** $R_{\hat{\Delta}_N}\{\cdot\}$ is computed based on N samples.

The attack signal must be stealthy for all sampled realizations. Very likely to be bounded!



Risk-optimal defense:

$$\min_{\Gamma} R_{\hat{\Delta}_N}(\Gamma)$$

Anand et al.. "Risk Assessment of Stealthy Attacks on Uncertain Control Systems". IEEE TAC, 2023

Gallo et al. "Design of multiplicative watermarking against covert attacks". CDC 2021

Outline

- Security Risk Management
- Scenario and Threat Models
- Security Metrics and Game-Theoretic Design
- Security under Model Uncertainty
- Probabilistic Risk Measures and Game-Theoretic Design
- **Conclusions and Remarks**

Conclusions and Remarks

- Risk management is a more comprehensive term than *security*.
- The importance of Adversary models to define (in)security.
- Security metrics - a bridge between risk management and game-theoretic design
- The role of Uncertainty and its relation to the Adversary.
    - Worst-case allows colluding with Adversary
    - Omniscient vs Bounded-Rationality
    - Uncertainty can be used as a form of defense (MTD)

Acknowledgments

Swedish Research Council

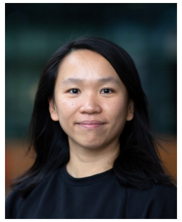
Swedish Foundation *for* Strategic Research

Uppsala Secure Learning and Control Lab

USLC

https://uslc-lab.github.io/

TU Delft



KTH VETENSKAP OCH KONST

UPPSALA UNIVERSITET

Acknowledgments

Swedish Research Council

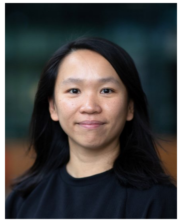
Swedish Foundation for Strategic Research

Uppsala Secure Learning and Control Lab

USLC

https://uslc-lab.github.io/

TU Delft



KTH VETENSKAP OCH KONST

UPPSALA UNIVERSITET

We are hiring!

USLC

Uppsala
Secure Learning
and Control Lab

https://uslc-lab.github.io/

Knut and Alice Wallenberg Foundation

Swedish Research Council

Swedish Foundation for Strategic Research

VINNOVA
Sweden's Innovation Agency

Swedish Energy Agency

UPPSALA
UNIVERSITET

We are hiring!

Uppsala
Secure Learning
and Control Lab

https://uslc-lab.github.io/

Swedish
Research Council

Swedish Foundation for Strategic Research

Sweden's Innovation Agency

Swedish Energy Agency

Openings:

- Hiring 1 Postdoc in Secure Federated Learning
- Hiring 1 PhD student in Distributed Voltage Control
- More positions to come in 2024/2025!

Uppsala Universitet

Backup slides

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

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

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)

Classical fault-tolerant control design objectives

Classical fault-tolerant control design objectives

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

Classical fault-tolerant control 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

Classical fault-tolerant control design objectives

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

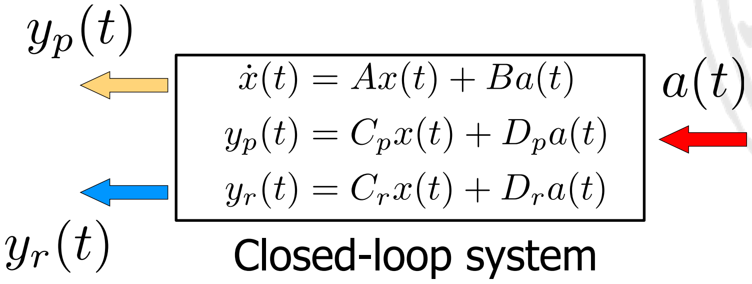
- **Fault detection filter design:** find an observer/filter that
  - Maximizes the “worst-case” (smallest) **detectability** of unit-energy faults
  - i.e.: optimal H_- detection filter design

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

Classical Sensitivity Metrics

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

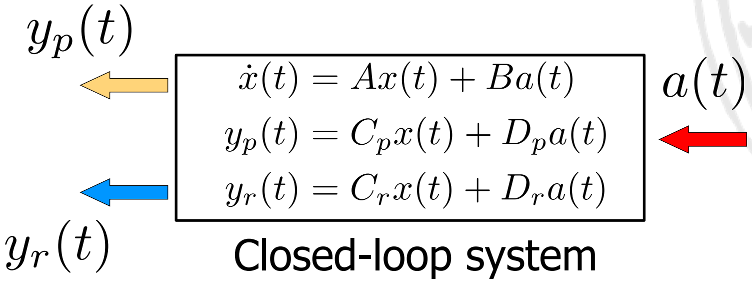
$$\|y\|_{\mathcal{L}_2}^2 \triangleq \int_{-\infty}^{+\infty} \|y(t)\|_2^2 \, dt$$



Classical Sensitivity Metrics

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

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



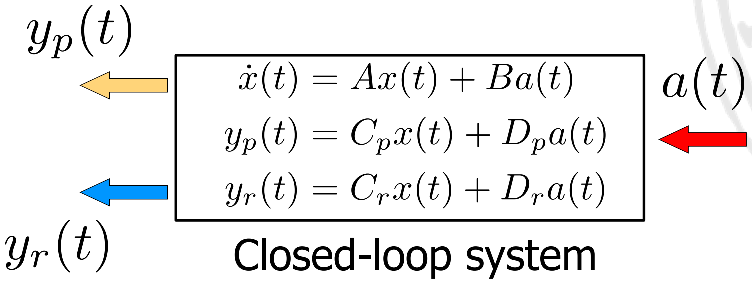
- Robustness:

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

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:

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

- Frequency Domain:

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

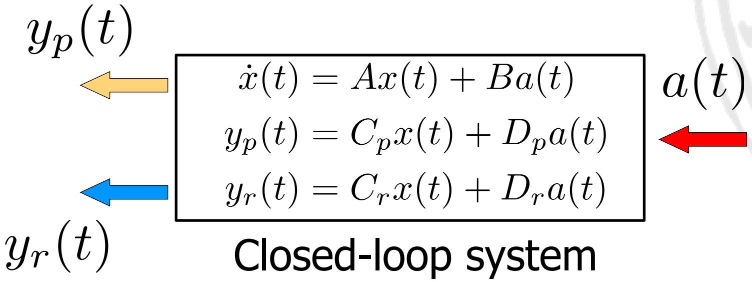
$$\bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^{n_a}} \|G_p(s)a\|_2 \quad \text{s.t.} \quad \|a\|_2 = 1$$

UPPSALA UNIVERSITET

Classical Sensitivity Metrics

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

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



- Robustness:

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

- Frequency Domain:

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

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

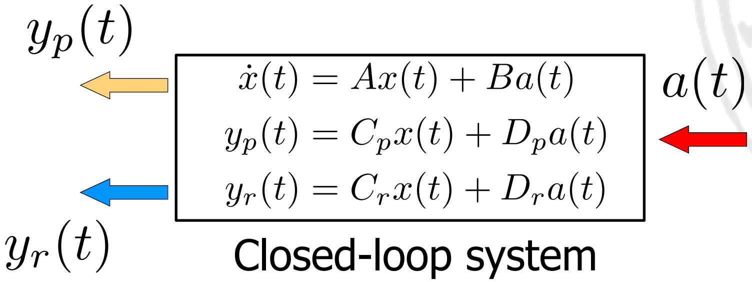
- Detectability:

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

Classical Sensitivity Metrics

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

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



- Robustness:

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

- Frequency Domain:

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

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

- Detectability:

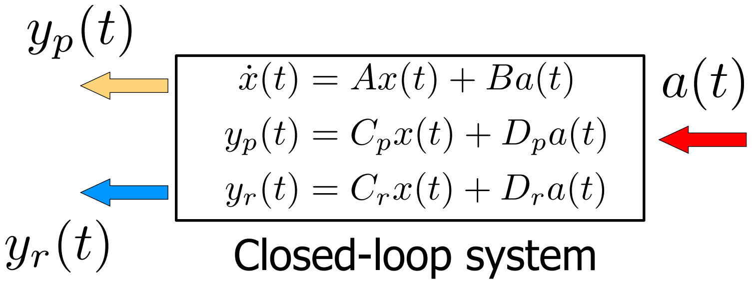
$$\gamma_{H_-} \triangleq \inf_{a \in \mathcal{L}_{2e}} \|y_r\|_{\mathcal{L}_2} \quad \text{s.t.} \quad \|a\|_{\mathcal{L}_2} = 1, \quad x(0) = 0$$

- Frequency Domain:

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

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

Classical Sensitivity Metrics

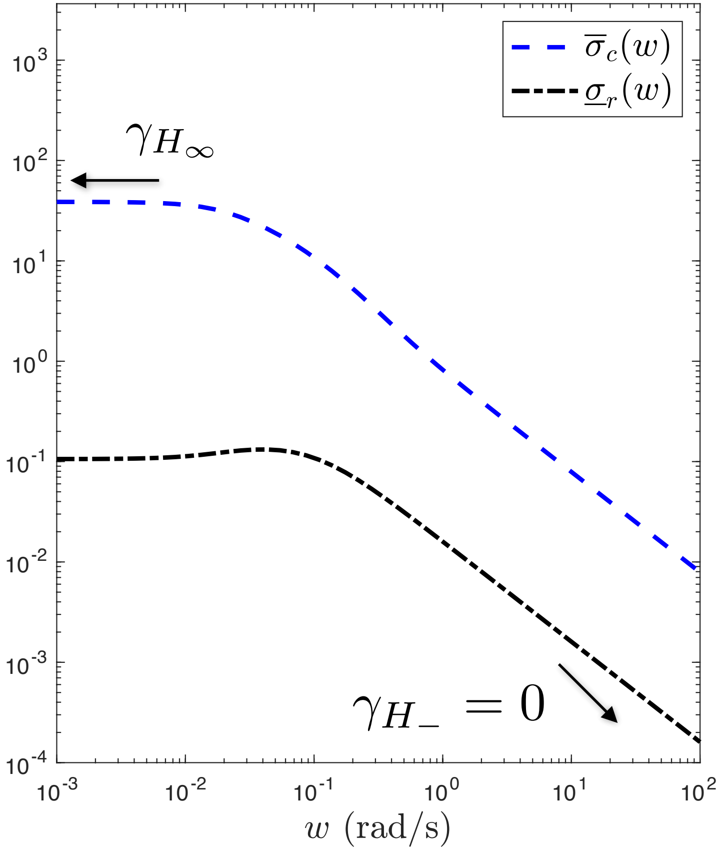


- Robustness (H_∞)

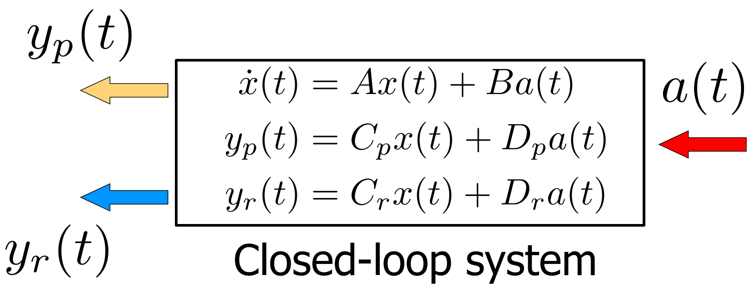
$$\gamma_{H_\infty} = \sup_{w \geq 0} \quad \bar{\sigma}_p(jw)$$

- Detectability ($H\_$)

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



Classical Sensitivity Metrics



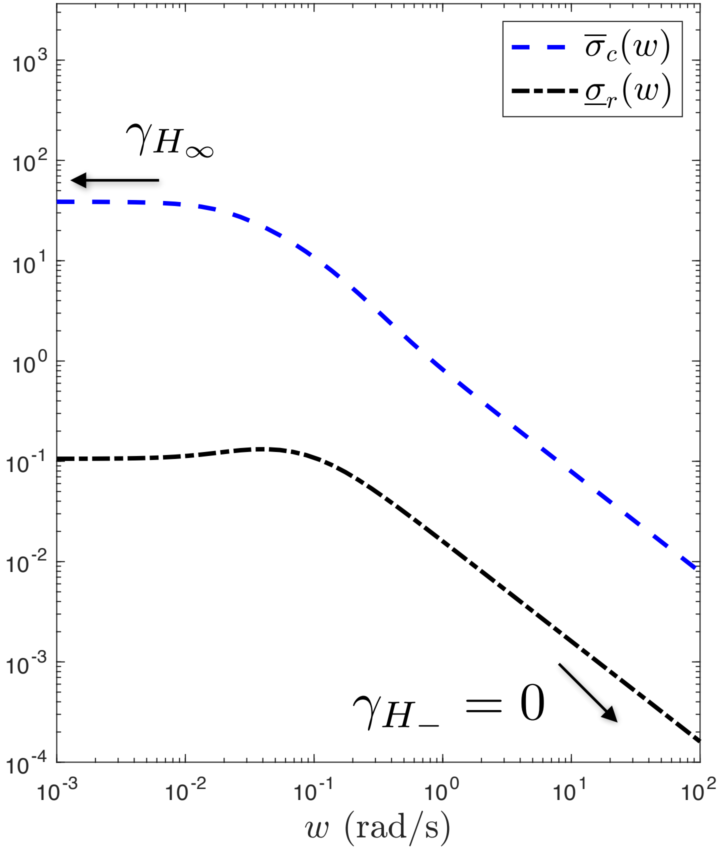
- Robustness (H_∞)

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

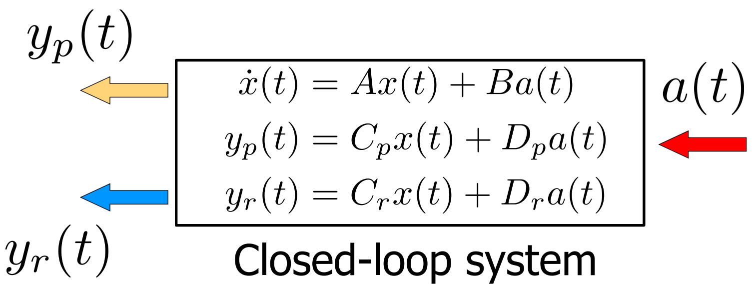
- Detectability (H_-)

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

- Least detectable fault has little impact…



Classical Sensitivity Metrics



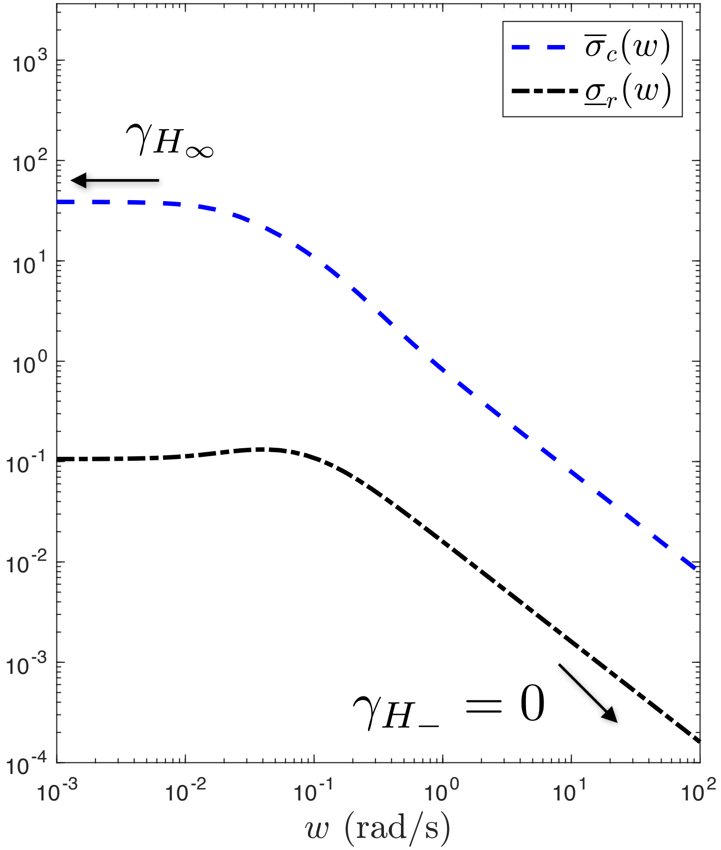
- Robustness (H_∞)

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

- Detectability ($H\_$)

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

- Least detectable fault has little impact…



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

Example: Robust Stealthy Attacks

Anand et al.. “Risk Assessment of Stealthy Attacks on Uncertain Control Systems”. IEEE TAC, 2023

SYSTEM PARAMETERS

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

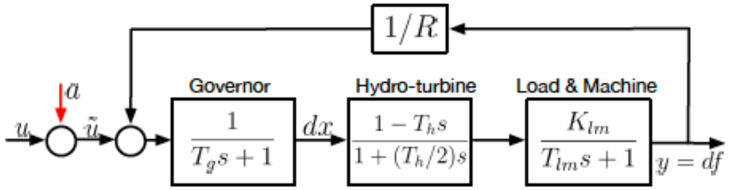


Fig. 4. Power generating system with a hydro turbine.

Gallo et al. “Design of multiplicative watermarking against covert attacks”. CDC 2021

UPPSALA UNIVERSITET

Example: Robust Stealthy Attacks

Anand et al.. "Risk Assessment of Stealthy Attacks on Uncertain Control Systems". IEEE TAC, 2023

SYSTEM PARAMETERS

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

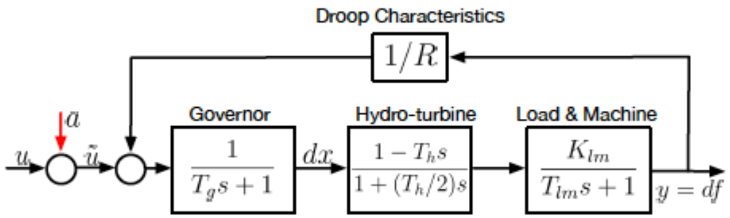


Fig. 4. Power generating system with a hydro turbine.

Without uncertainty:

- Unbounded impact for any parameter value $T_h \in [4,6]$.

Gallo et al. "Design of multiplicative watermarking against covert attacks". CDC 2021

UPPSALA UNIVERSITET

Example: Robust Stealthy Attacks

Anand et al.. “Risk Assessment of Stealthy Attacks on Uncertain Control Systems”. IEEE TAC, 2023

System Parameters

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

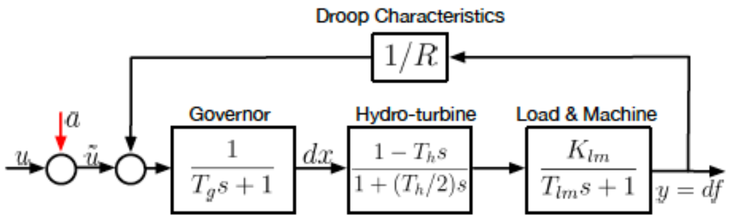


Fig. 4. Power generating system with a hydro turbine.

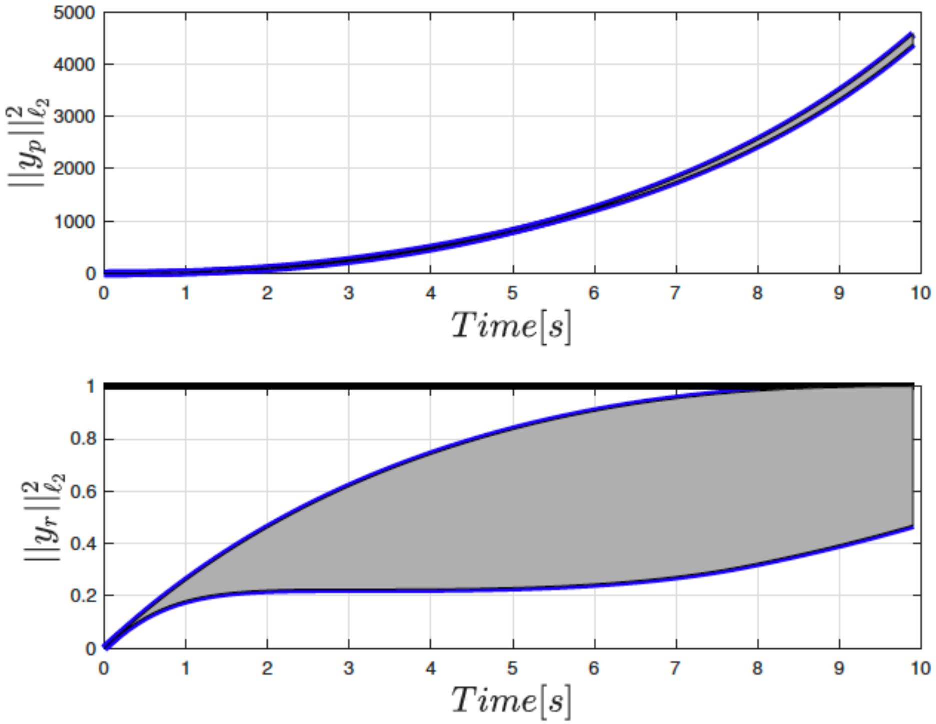
Without uncertainty:

- Unbounded impact for any parameter value $T_h \in [4{,}6]$.

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, …

Gallo et al. “Design of multiplicative watermarking against covert attacks”. CDC 2021



UPPSALA UNIVERSITET

References

- [Anand and Teixeira, IFAC 2020] S.C. Anand and **A. M. H. Teixeira**. “Joint controller and detector design against data injection attacks on actuators”. In Proc. IFAC World Congress, Berlin, Germany, Jul. 2020.
- [Cardenas *et al*., HOTSEC, 2008] A.A. Cárdenas, S. Amin, S. Sastry, “Research Challenges for the Security of Control Systems”, Proceedings of the 3rd conference on Hot topics in Security, HotSec 2008
- [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.
- [Ferrari and Teixeira, TAC 2020] R. Ferrari, **A. M. H. Teixeira**. “A Switching Multiplicative Watermarking Scheme for Detection of Stealthy Cyber-Attacks”. IEEE Transactions on Automatic Control, 2020. Early Access.
- [Kaplan & Garrick, 1981] S. Kaplan and B. J. Garrick, “On The Quantitative Definition of Risk”, Risk Analysis, vol. 1, issue 1, 11-27, 1981.
- [Milosevic *et al*., TAC 2020] J. Milosevic, **A.M.H. Teixeira**, H. Sandberg, K. H. Johansson. “Actuator Security Indices Based on Perfect Undetectability: Computation, Robustness, and Sensor Placement”. IEEE Transactions on Automatic Control, vol. 65, no. 9, pp. 3816–3831, 2020.
- [Milosevic *et al*., IJRNC 2018] J. Milosevic, **A. M. H. Teixeira**, T. Tanaka, K. H. Johansson, H. Sandberg. “Security Measure Allocation for Industrial Control Systems: Exploiting Systematic Search Techniques and Submodularity”. International Journal of Robust and Nonlinear Control 3(1):4278-4302, Jul. 2020
- [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.
- [Mo and Sinopoli, CPS Week 2010] False data injection attacks in control systems. Yilin Mo and Bruno Sinopoli. First Workshop on Secure Control Systems, CPS Week, 2010
- [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.
- [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.
- [Tang *et al.,* Automatica 2019] Z. Tang, M. Kuijper, M. S. Chong, I. Mareels, C. Leckie, "Linear system security - Detection and correction of adversarial sensor attacks in the noise-free case”, Automarica, 101:53-59, 2019.
- [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.
- [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
- [Teixeira, CDC 19] **A.M.H. Teixeira**, “Optimal stealthy attacks on actuators for strictly proper systems”, *2019 58th IEEE Conference on Decision and Control (CDC)*, Nice, France, Dec 2019
- [Teixeira, Springer 2021] **A.M.H. Teixeira**, “Security Metrics for Control Systems”, in Safety, Security, and Privacy for Cyber-Physical Systems, R. M. G. Ferrari and A. M. H. Teixeira, Eds. Springer International Publishing. In production.
- [“The Real Story of Stuxnet”, IEEE Spectrum, 2013] D. Kushner, “The Real Story of Stuxnet”, IEEE Spectrum, vol. 50, no. 3, pp. 48-53, 2013.

Additional References

- [W72] 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
- [HM80] David J. Hill, Peter J. Moylan, “Dissipative Dynamical Systems: Basic Input-Output and State Properties”, Journal of the Franklin Institute, 309 (5) (1980), pp. 327-357.
- [TW91] 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)
- [W71] J. C. Willems, “Least Squares Stationary Optimal Control and the Algebraic Riccati Equation,” IEEE Transactions on Automatic Control, 16 (6) (1971), pp. 621-634.
- Hamayun M.T., Edwards C., Alwi H. (2016) Fault Tolerant Control. In: Fault Tolerant Control Schemes Using Integral Sliding Modes. Studies in Systems, Decision and Control, vol 61. Springer, Cham