KTH ROYAL INSTITUTE
OF TECHNOLOGY

A Risk Management Approach to Cyber-Physical Security in Networked Control Systems

Henrik Sandberg
Decision and Control Systems, KTH EECS, hsan@kth.se

Outline

- Motivation and background

- Part 1: Risk Management
  - Scenario characterization
  - Risk analysis
  - Risk mitigation
  - Examples

- Part 2 (time permitting): Minimum-time Secure Rollout of Software Updates for Controllable Power Loads

- Summary and outlook

References

- Part 1:

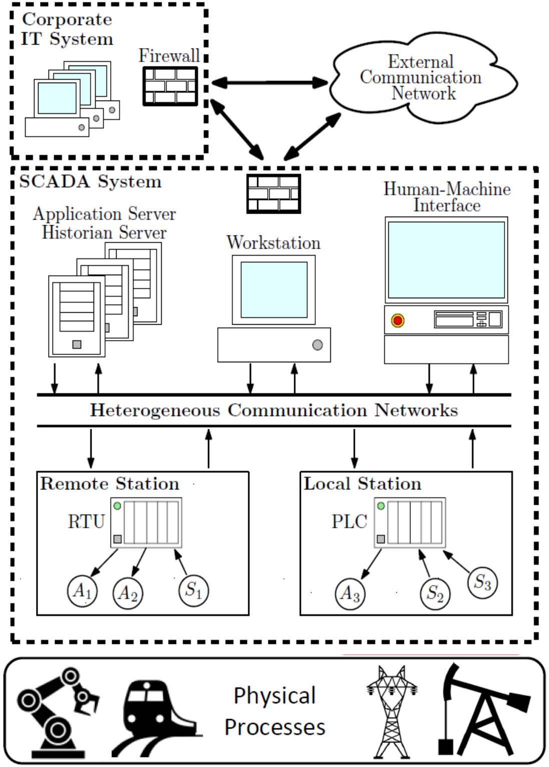
[1] M.S. Chong, H. Sandberg, A.M.H. Teixeira: “A Tutorial Introduction to Security and Privacy for Cyber-Physical Systems”. 18th European Control Conference (ECC), Naples, Italy, 2019, pp. 968-978.

- Part 2:

[2] M.G. de Medeiros, Kin Cheong Sou, Henrik Sandberg: “Minimum-time Secure Rollout of Software Updates for Controllable Power Loads”. Electric Power Systems Research, 189, 106797, Dec 2020.

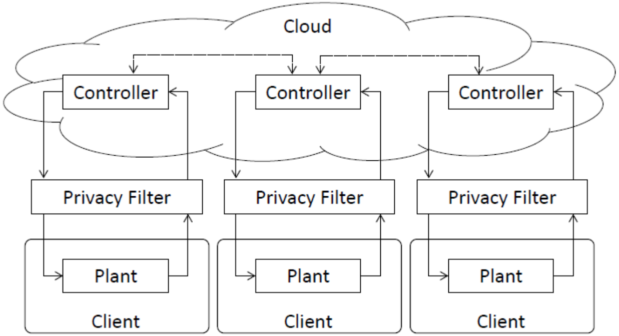
Cyber-Physical Systems

Industrial Control System (ICS)



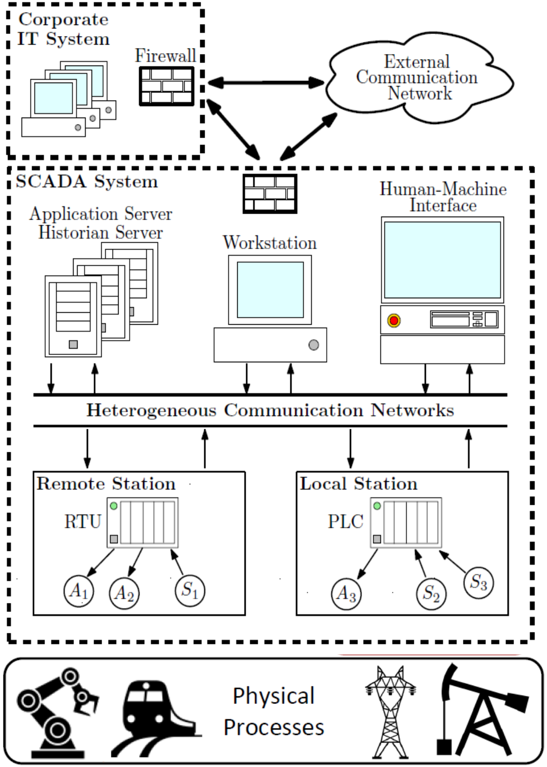
Autonomous vehicles

Cloud-based Control and IoT



Typical ICS Vulnerabilities

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



Example 1: The Stuxnet Worm (2010)

Targets: Windows, ICS, and PLCs connected to variable-frequency drives

Exploited **4 zero-day flaws**

- **Goal:**

Harm centrifuges at uranium enrichment facility in Iran

- **Attack mode:**

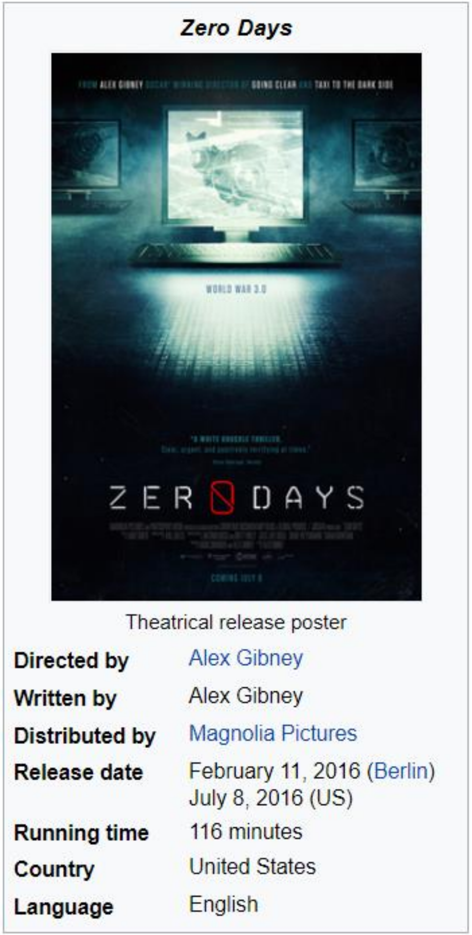
1. Delivery with USB stick (**no internet connection necessary**)
2. Replay measurements to control center and execute harmful controls

["The Real Story of Stuxnet", IEEE Spectrum, 2013]

Example 1: Stuxnet (2010)

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

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



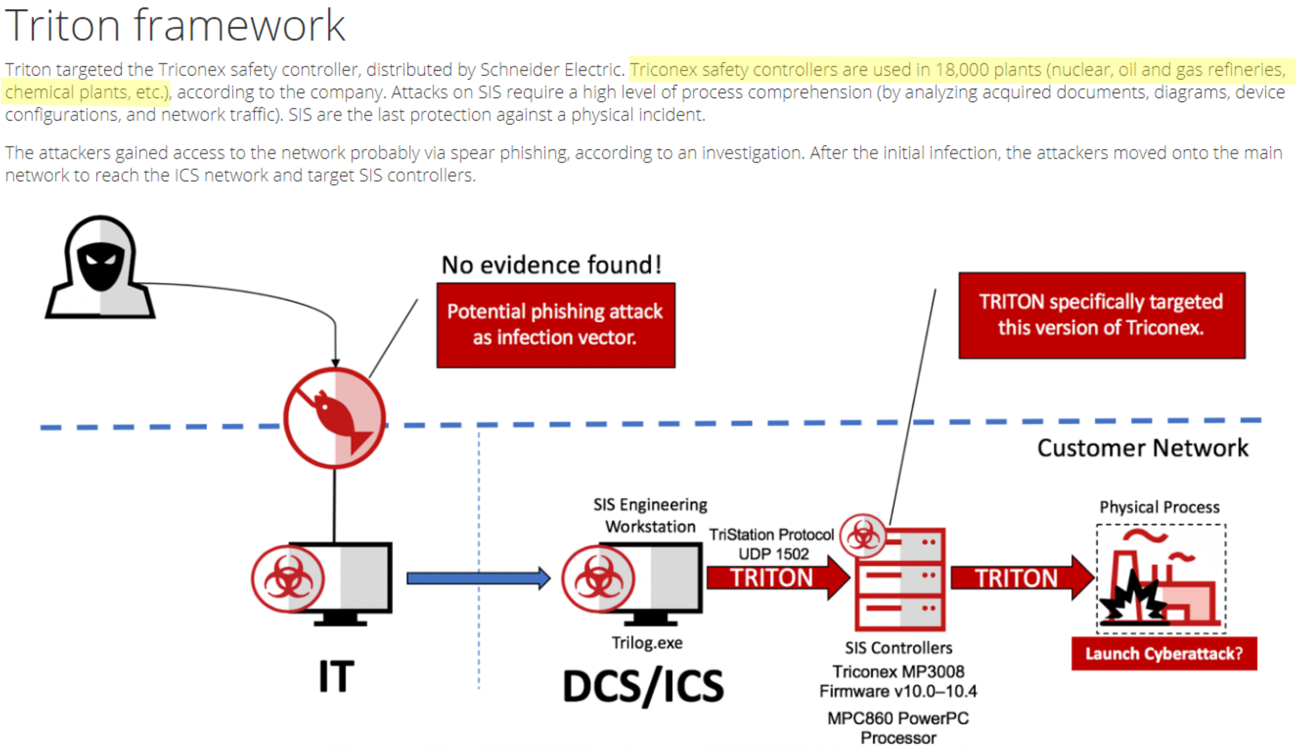
Zero Days

Theatrical release poster

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

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

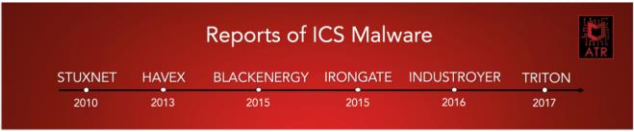
Example 2: Triton Malware (2017)



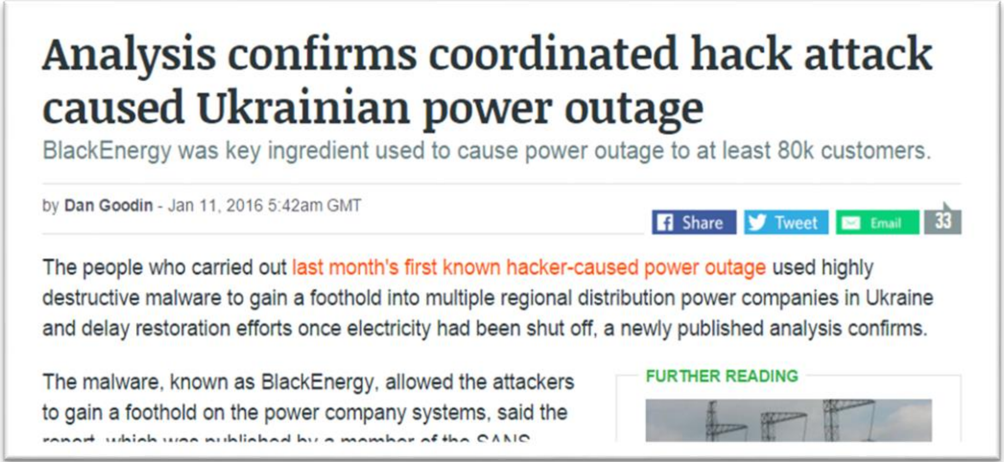
Triton framework

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

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



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



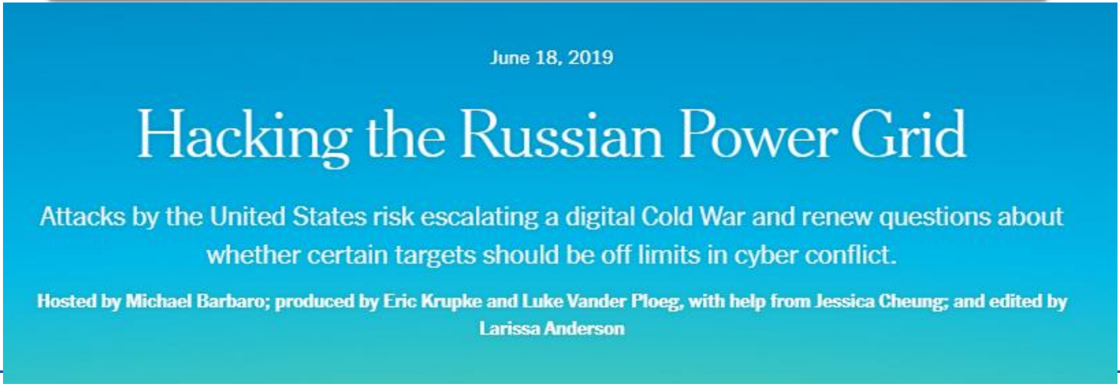
Analysis confirms coordinated hack attack caused Ukrainian power outage

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

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

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

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



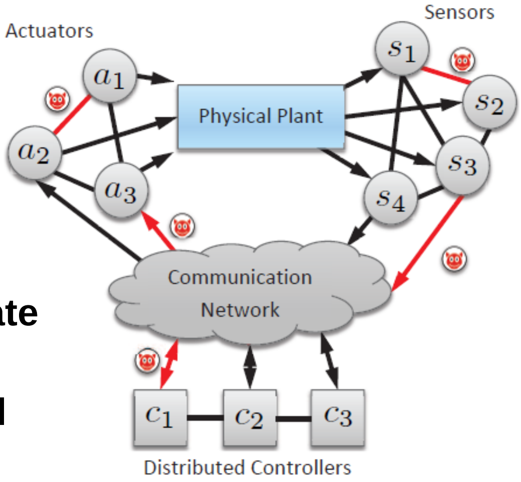
June 18, 2019

Hacking the Russian Power Grid

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

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

Cyber-Secure Control



Networked control systems

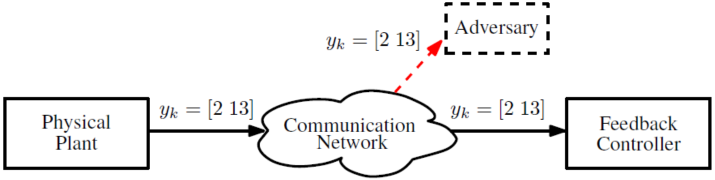
- are being **integrated with business/corporate networks**
- have many potential points of **cyber-physical attack**

Need tools and strategies to understand and mitigate attacks:

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

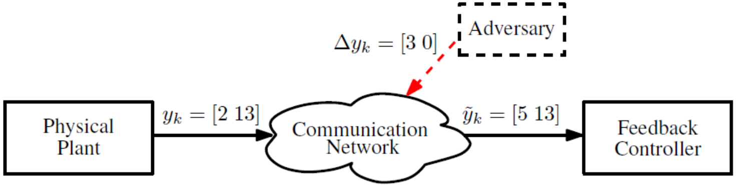
CIA in IT Security [Bishop, 2002]

- **C – Confidentiality**
  - [1]: See work by Le Ny, for example



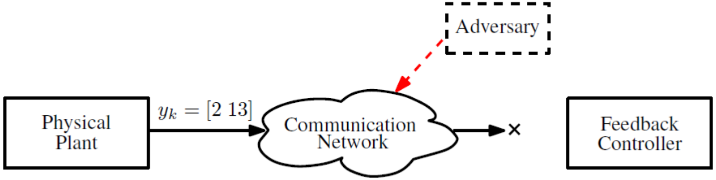
(a) Data confidentiality violation by a disclosure attack.

- **I – Integrity**
  - [1]: See work by Tabuada, Sandberg, Sinopoli, for example



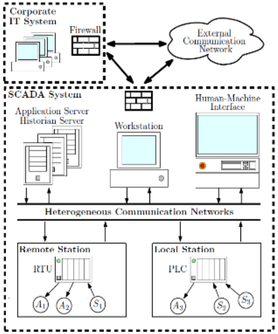
(b) Data integrity violation by a false-data injection attack.

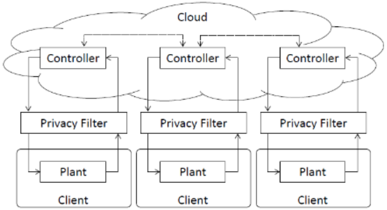
- **A – Availability**
  - [1]: See work by Tesi, for example



(c) Data availability violation by a denial-of-service attack.

Is More Than IT Security and Fault Tolerance Needed?



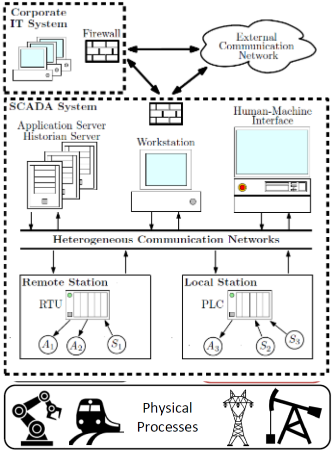


- **Clearly IT security and fault tolerance are needed**: Authentication, encryption, firewalls, error correction, etc.

But not sufficient…

- **Interaction between physical and cyber systems** make control systems different from normal IT systems
- **Malicious actions can enter anywhere** in the closed loop and cause harm, whether channels secured or not
- **Malicious attackers** have an **intent,** as opposed to faults, and can act strategically
- **Can we trust** the interfaces and channels are really secured? (see **OpenSSL** Heartbleed bug…)

Security Challenges in ICS



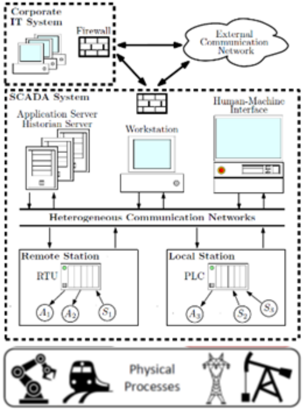
Differences to traditional IT systems:

- **Patching and frequent updates are not well suited for control systems** (see Part 2)
- **Real-time availability** (Strict operational environment)
- **Legacy systems** (Often no authentication or encryption)
- **Protection of information and physical world** (Estimation and control algorithms)

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

[Cardenas *et al*., HOTSEC, 2008]

Security Challenges in ICS



“New” vulnerabilities and “new” threats:

- Controllers are computers (Relays → Microprocessors)
- Networked (Access from corporate network)
- Commodity IT solutions (Windows, TCP/IP,…)
- Open design (Protocols known)
- Increasing size and functionality (New services, wireless,...)
- Large and highly skilled IT global workforce (More IT knowledge)
- Cybercrime (Attack tools available)

[Cardenas *et al*., HOTSEC, 2008]

Outline

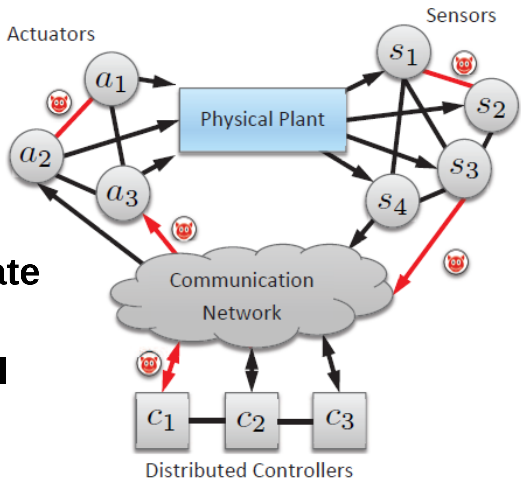
- Motivation and background

- **Part 1: Risk Management**
  - Scenario characterization
  - Risk analysis
  - Risk mitigation
  - Examples

- Part 2 (time permitting): Minimum-time Secure Rollout of Software Updates for Controllable Power Loads

- Summary and outlook

Cyber-Secure Control



Networked control systems

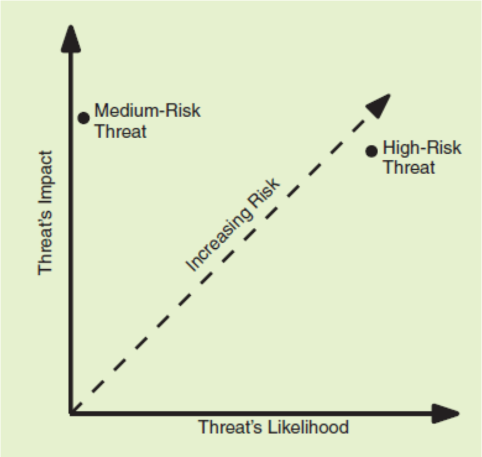
- are being **integrated with business/corporate networks**
- have many potential points of **cyber-physical attack**

Need tools and strategies to understand and mitigate attacks:

- **Which threats** should we care about?
- **What impact** can we expect from attacks?
- **Which resources** should we **protect** (more), and how?
- **Answer: Risk management**

Defining Risk

Risk = (Scenario, Likelihood, Impact)



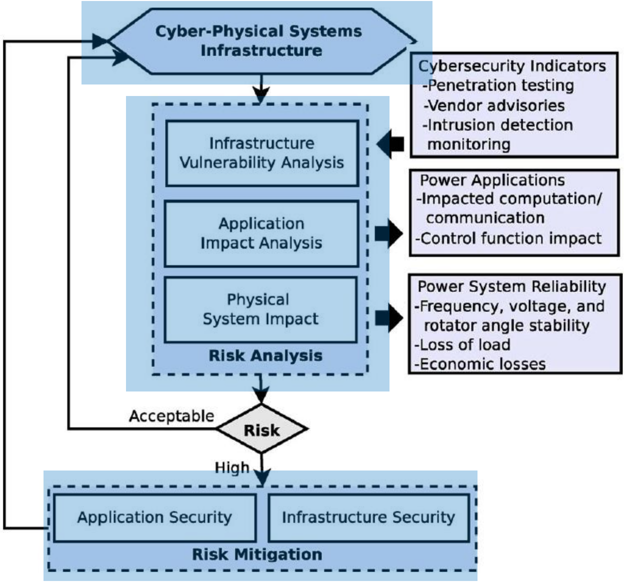
- **Scenario**
  - How to describe the system under attack?
- **Likelihood**
  - Interpretations:
    - *a) Likelihood of attack in progress being successful (experts' assessment)*
    - *b) Likelihood = 1*
    - *c) ~1/effort to conduct attack*
- **Impact**
  - What are the cyber-physical consequences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

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

Risk Management Cycle

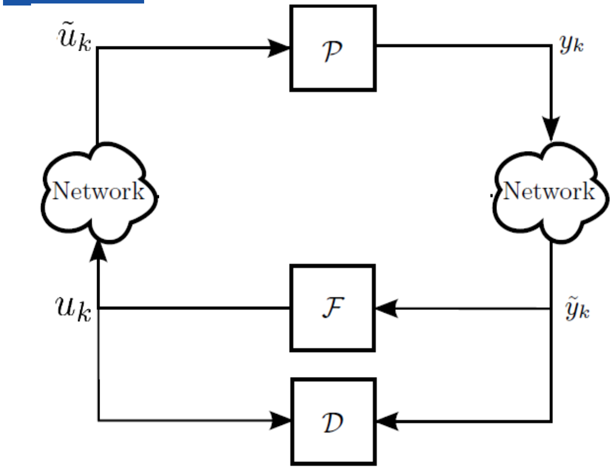
Main steps in risk management

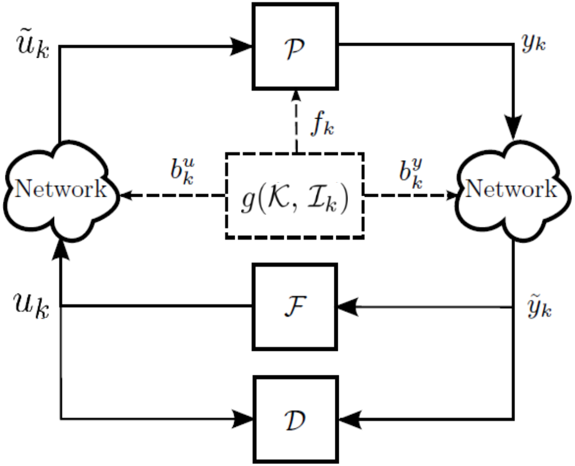
- Scenario characterization
  - Models, Scenarios, Objectives
- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment
- Risk Mitigation
  - Prevention, Detection, Treatment



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

Networked Control System under Attack





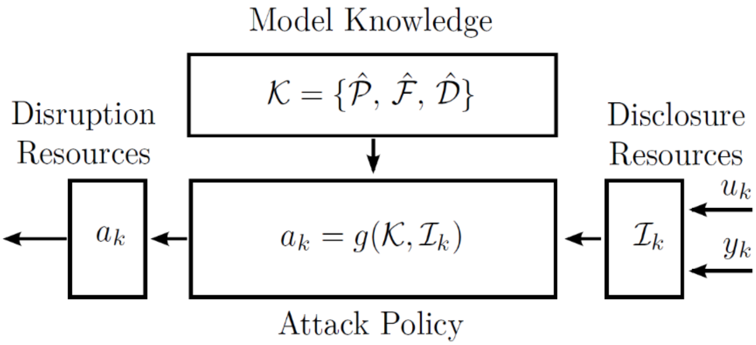
- Physical plant ($\mathcal{P}$)
- Feedback controller ($\mathcal{F}$)
- Anomaly detector ($\mathcal{D}$)
- Disclosure Attacks

- Physical Attacks f_k
- Data Injection Attacks

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

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

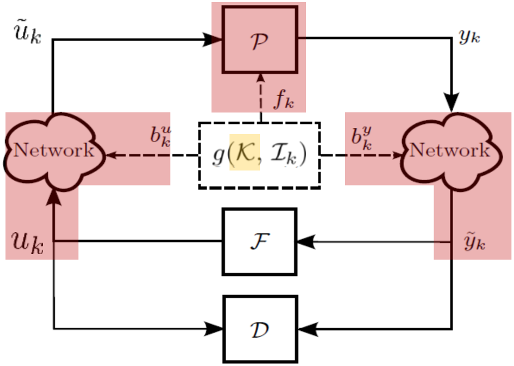
Adversary Model

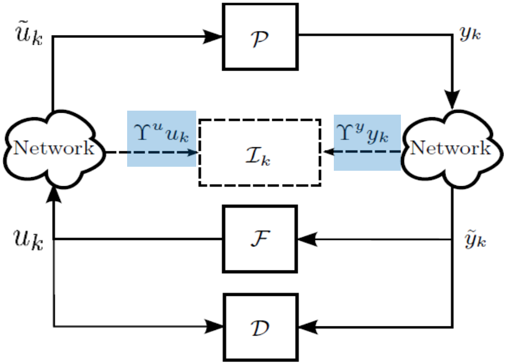

Model Knowledge
$\mathcal{K} = \{\hat{\mathcal{P}}, \hat{\mathcal{F}}, \hat{\mathcal{D}}\}$
Disruption Resources
Disclosure Resources
a_k
$a_k = g(\mathcal{K}, \mathcal{I}_k)$
$\mathcal{I}_k$
u_k
y_k
Attack Policy

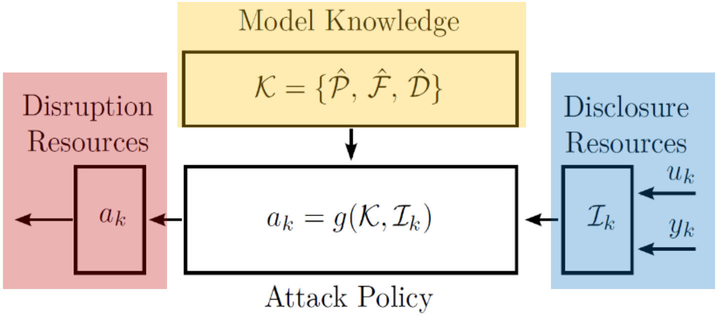
- **Attack policy:** Goal of the attack? Destroy equipment, increase costs, *remain undetected*…
- **CPS model knowledge:** Adversary knows models of plant and controller? Possibility for stealthy attacks…
- **Disruption/disclosure resources:** Which channels can the adversary access?

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

Networked Control System with Adversary Model

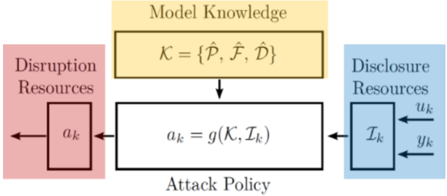


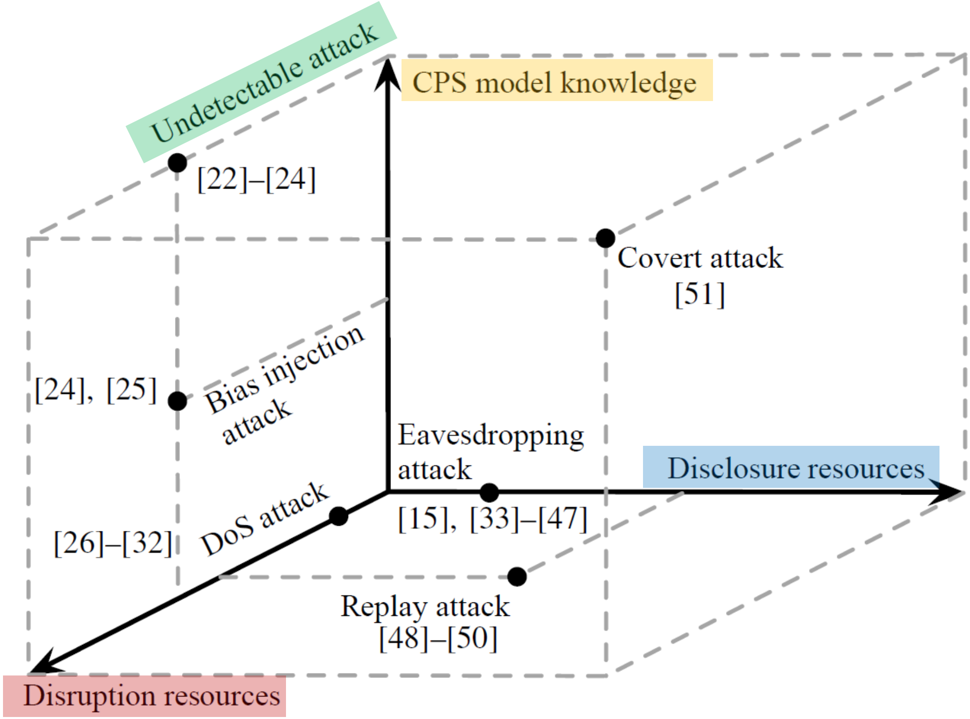




Model Knowledge
$\mathcal{K} = \{\hat{\mathcal{P}}, \hat{\mathcal{F}}, \hat{\mathcal{D}}\}$
Disruption Resources
a_k
$a_k = g(\mathcal{K}, \mathcal{I}_k)$
Disclosure Resources
$\mathcal{I}_k$
u_k
y_k
Attack Policy

CPS Attack Space [1]


Model Knowledge
$\mathcal{K} = \{\hat{\mathcal{P}}, \hat{\mathcal{F}}, \hat{\mathcal{D}}\}$
Disruption Resources
a_k
$a_k = g(\mathcal{K}, \mathcal{I}_k)$
Disclosure Resources
u_k
$\mathcal{I}_k$
y_k
Attack Policy

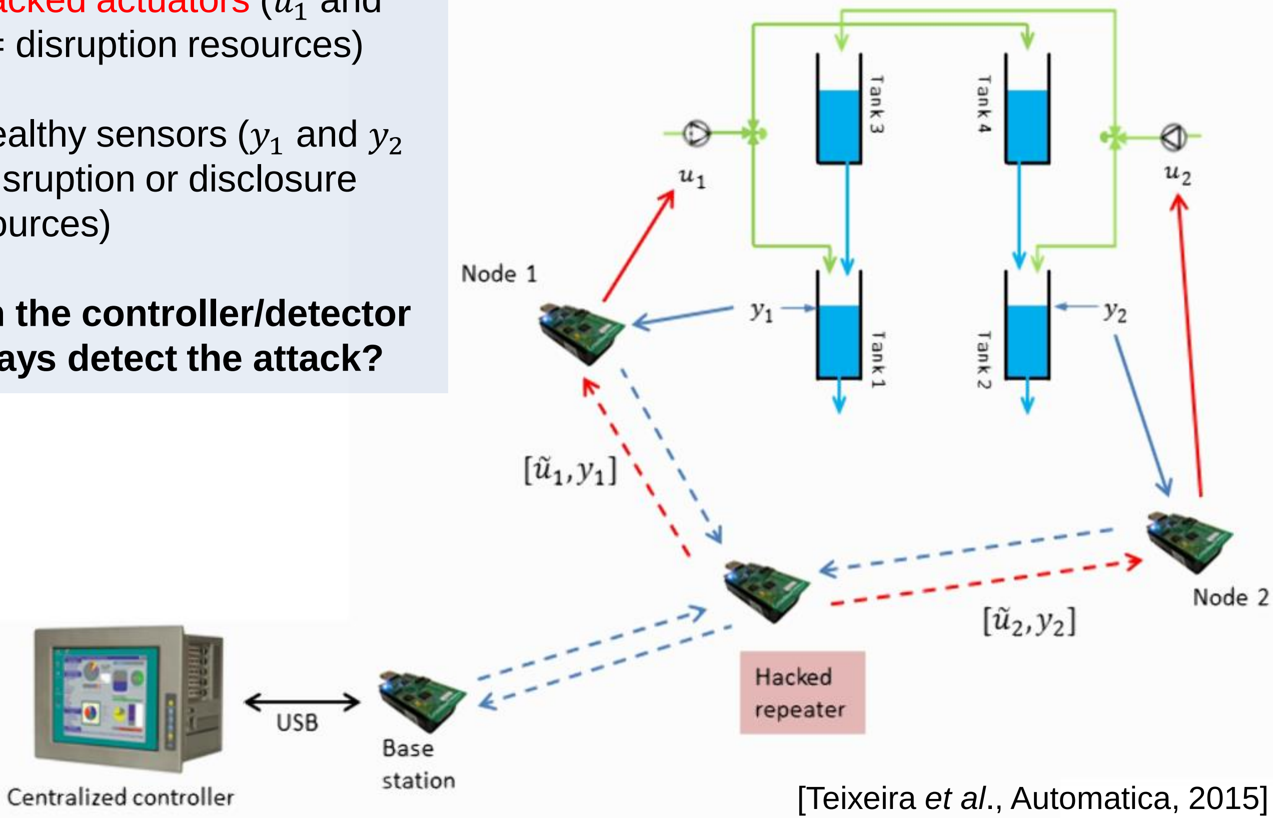


Example: Undetectable Water Tank Attack

2 hacked actuators (u_1 and u_2 = disruption resources)

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

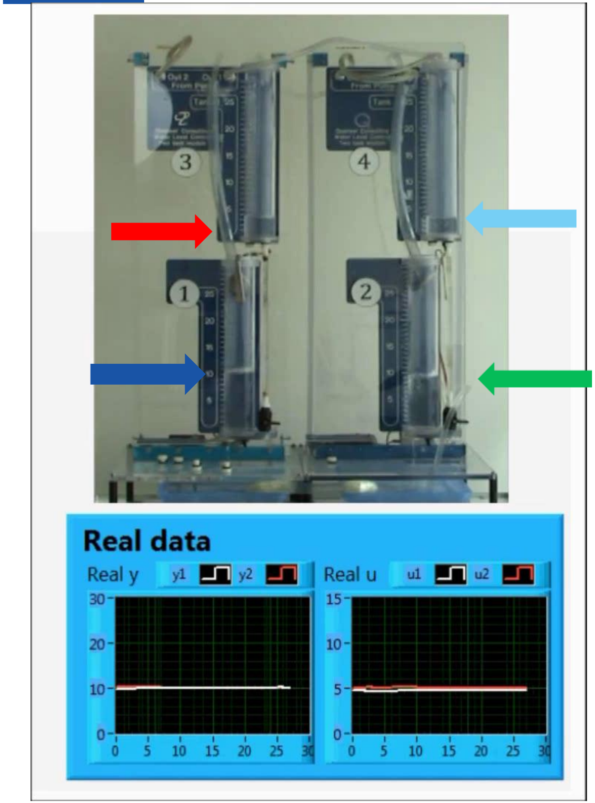
Can the controller/detector always detect the attack?

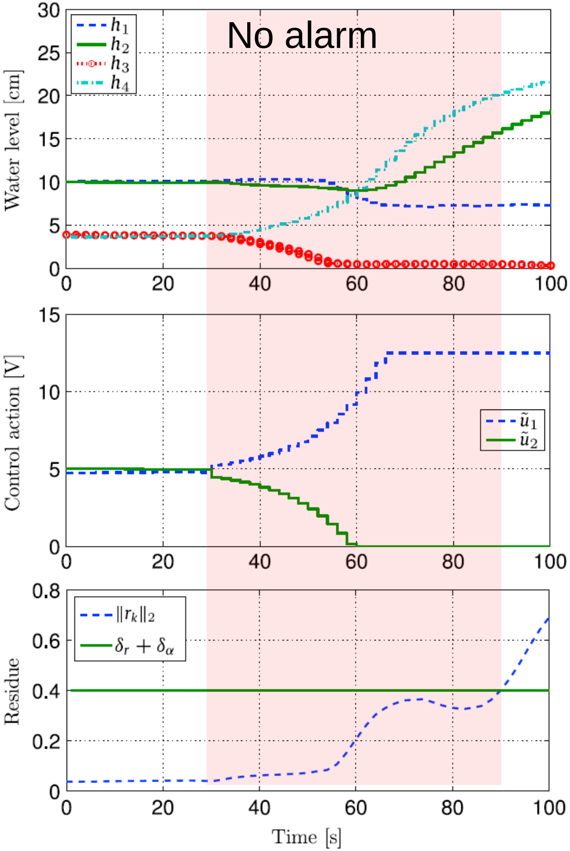


[Teixeira *et al*., Automatica, 2015]

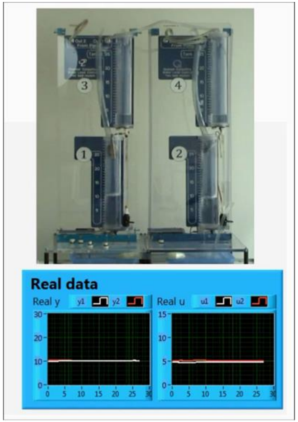
[Teixeira *et al.*, Automatica, 2015]

Undetectable Water Tank Attack [Movie]





Water Tank Model Analysis



- Transfer function matrix from attack to sensor signals

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

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

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

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

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

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

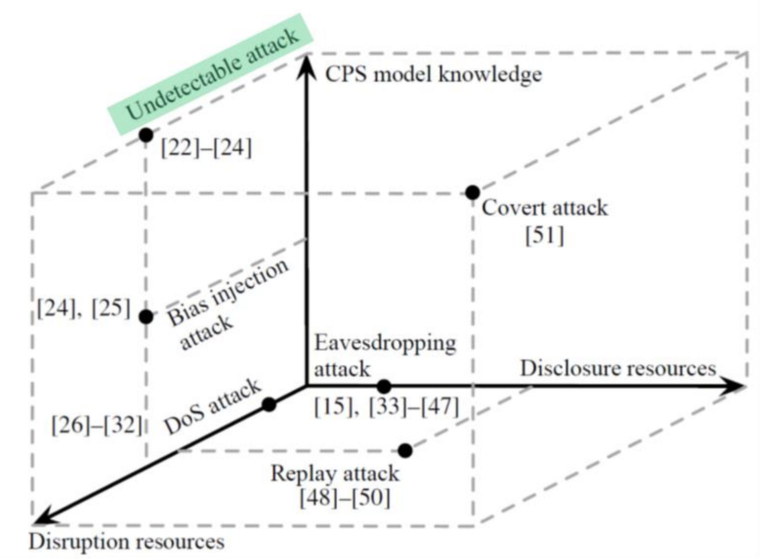
Undetectable Water Tank Attack

2 hacked actuators (u_1 and u_2 = disruption resources)

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

Can the controller/detector always detect the attack?

Not against an adversary with CPS model knowledge



Undetectable Attacks: General Case

- Consider the linear system $y = G_d d + G_a a$ (the controlled infrastructure):

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

- **Operator:** State $x(k) \in \mathbb{R}^n$, disturbance $d(k) \in \mathbb{R}^o$, and (malicious) attack $a(k) \in \mathbb{R}^m$ *unknown*. Measurement $y(k) \in \mathbb{R}^p$, and model A, B_d, B_a, C, D_d, D_a *known*

- **Attacker:** Model A, B_d, B_a, C, D_d, D_a *known,* and can change attack $a(k) \in \mathbb{R}^m$ arbitrarily

Definition: Attack signal a is *undetectable* if there exists a simultaneous (masking) disturbance signal d and initial state $x(0)$ such that $y(k) = 0$, $k \geq 0$

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

Remark: Less strict stealthy attacks ($y(k) \approx 0$) defined in deterministic [Teixeira *et al*., Automatica, 2015] and stochastic [Bai *et al*., Automatica, 2017] setting

Undetectable Attacks and Invariant Zeros

- The Rosenbrock system matrix:

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

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

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

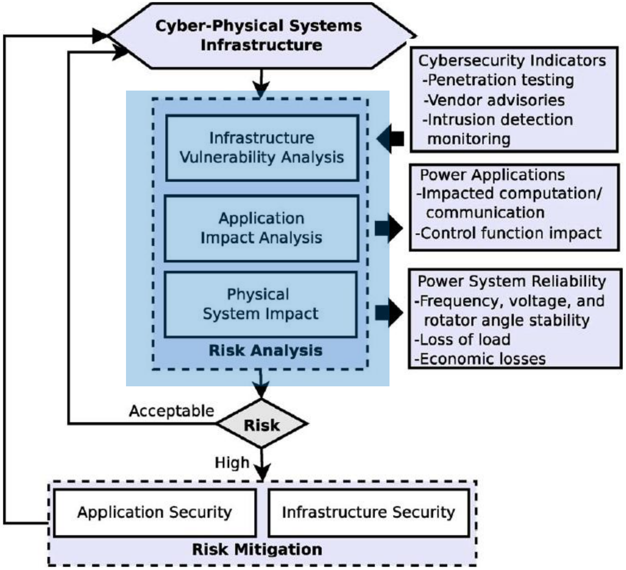
- Routine invariant zero computation **(**MATLAB: tzero**)**

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

Risk Management Cycle

Main steps in risk management

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



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

Tools for Likelihood Assessment: Security Index

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

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

Interpretation:

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

[Sandberg and Teixeira, SoSCYPS, 2016]

Tools for Likelihood Assessment: Security Index

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

Theorem 2: Suppose that the attacker can manipulate at most q components simultaneously ($\|a(k)\|_0 \leq q, \forall k$).

i. There exists persistent undetectable attacks $a^i \Leftrightarrow \; q \geq \alpha_i$

ii. All persistent attacks are i-identifiable $\Leftrightarrow q \; < \alpha_i/2$

iii. All persistent attacks are identifiable $\Leftrightarrow q \; < \min_i \alpha_i/2$

(See definition of identifiable next slide)

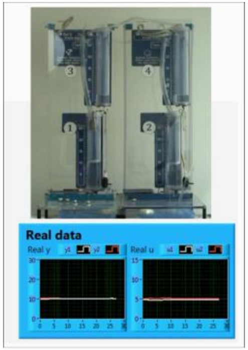
[Sandberg and Teixeira, SoSCYPS, 2016], see also [Tang *et al*., Automatica, 2019]

Attack Identification

- **Definition:** A (persistent) attack signal a is
- *identifiable* if for all attack signals $\tilde{a} \neq a$, and all corresponding disturbances d and $\tilde{d}$, and initial states $x(0)$ and $\tilde{x}(0)$, we have $\tilde{y} \neq y$;
- i*-identifiable* if for all attack signals a and $\tilde{a}$ with $\widetilde{a_i} \neq a_i$, and all corresponding disturbances d and $\tilde{d}$, and initial states $x(0)$ and $\tilde{x}(0)$, we have $\tilde{y} \neq y$

- **Interpretations:**
- Identifiability ⇔ (different attack a ⇒ different measurement y) ⇔ attack signal is injectively mapped to y ⇒ attack signal is detectable
- i*-identifiable* weaker than *identifiable*
- $\forall i$: a *is* i*-identifiable* ⇔ a is *identifiable*
- a is i*-identifiable:* Possible to track element a_i, but not necessarily a_j, $j \neq i$

Security Index Water Tank Example



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

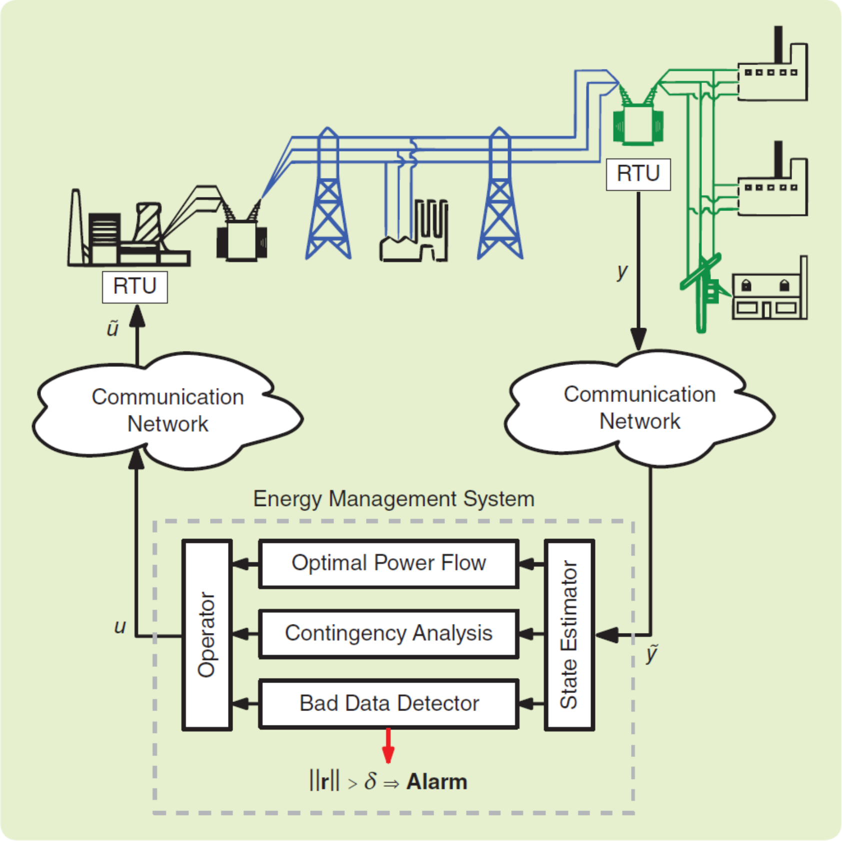
- Invariant zeros = $\{0.8675, 1.0362\}$ ⇒ Non-minimum phase system

- Persistent undetectable attack:

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

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

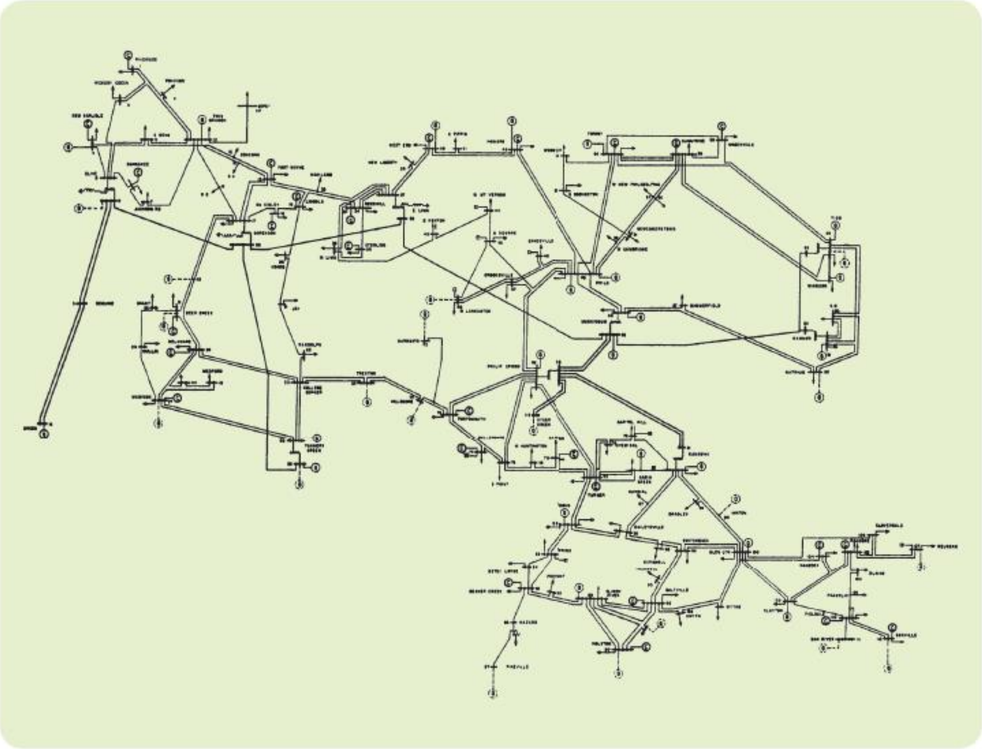
Example: Power System State Estimator



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

- State dimension $n = 118$

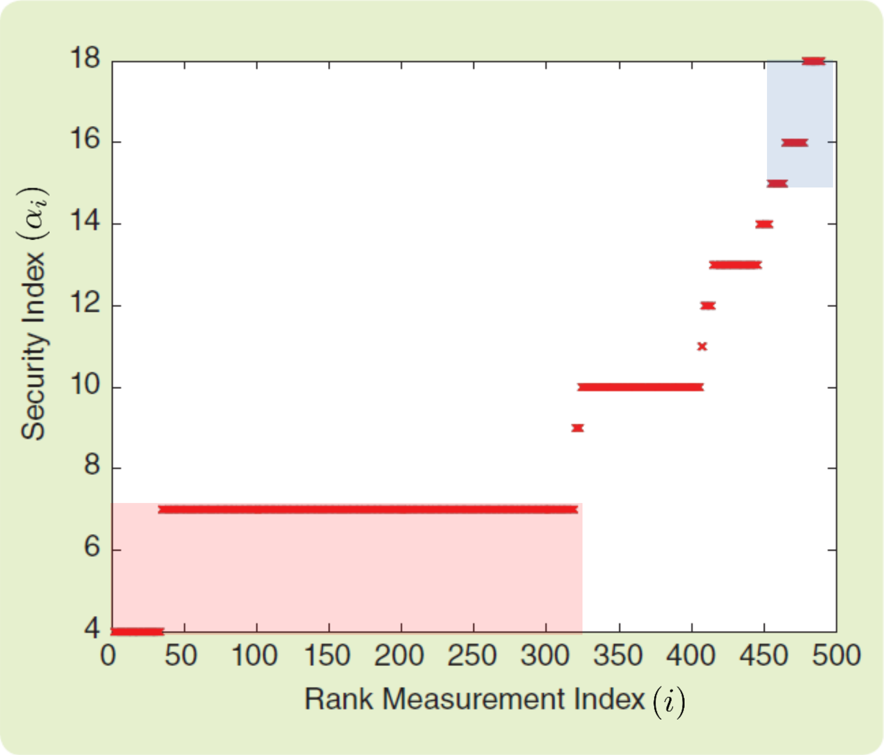
- Number sensors $p \approx 490$



[Teixeira *et al.*, IEEE CSM, 2015]

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

- Suppose number of attacked elements is $q \leq 7$. Theorem 2 yields:

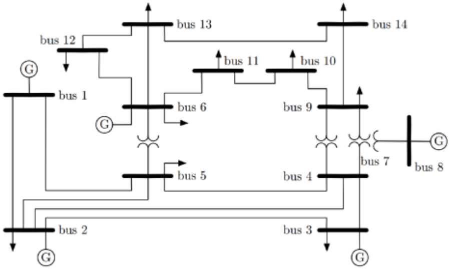


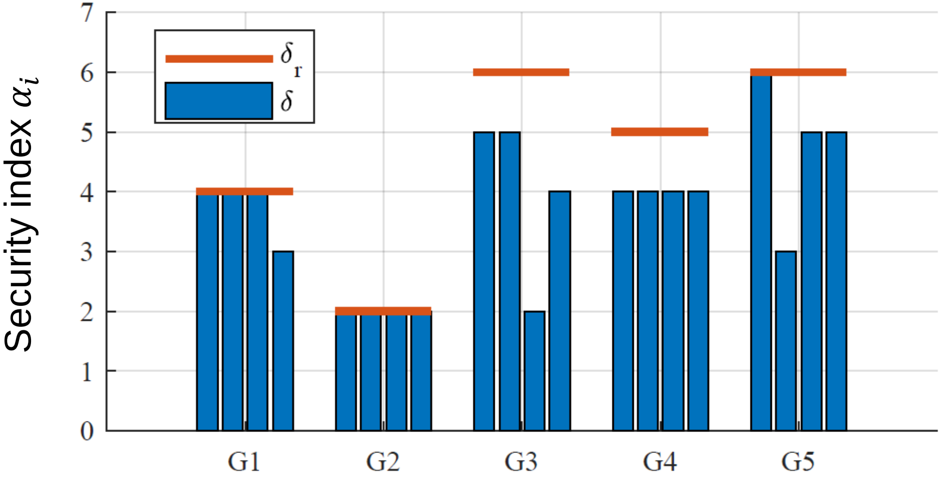
- Signals susceptible to undetectable attacks

- Signals where all attacks are identifiable

- Other signals will, if attacked, always result in non-zero output y

Example: Dynamic Security Index IEEE 14 Bus System



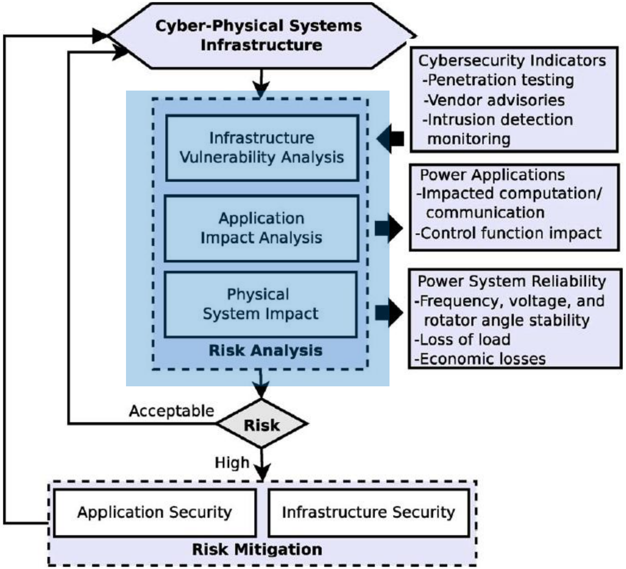


[Milošević *et al.*, IEEE TAC, 2020]

Risk Management Cycle

Main steps in risk management

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



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

Tools for Impact Assessment: Constrained Reachability

System model:

$$\begin{aligned} x(k+1) &= Ax(k) + B_a a(k) \\ y(k) &= Cx(k) + D_a a(k) \end{aligned}$$

Impact assessment problem:

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

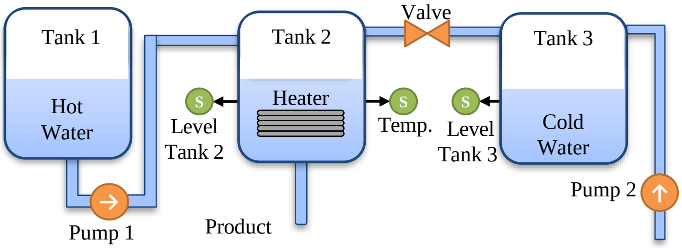
Theorem 3:

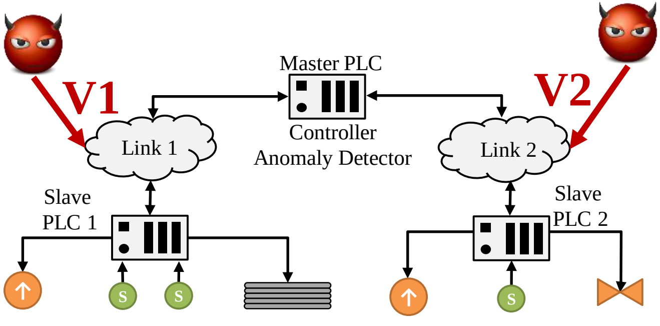
i. Constraints are convex

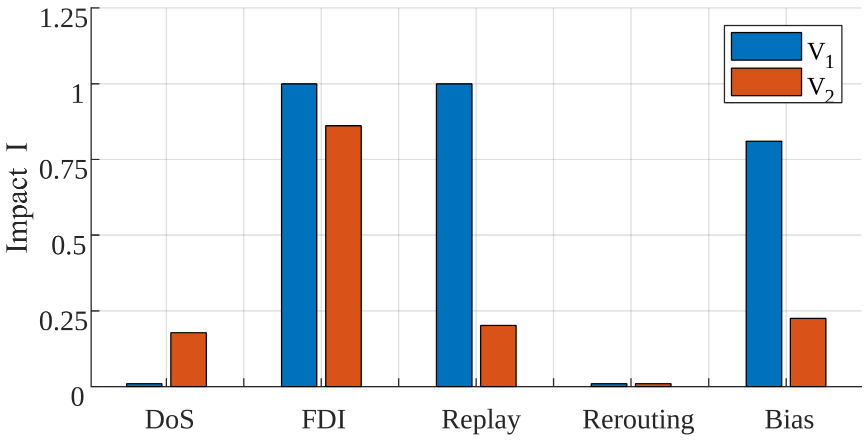
ii. Optimal value found by solving set of convex optimization problems

[Milošević *et al.*, ECC, 2018]

Tools for Impact Assessment: Constrained Reachability



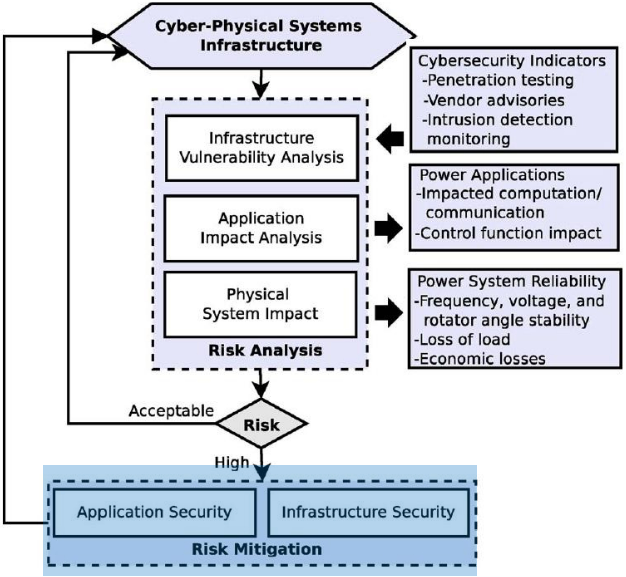




Risk Management Cycle

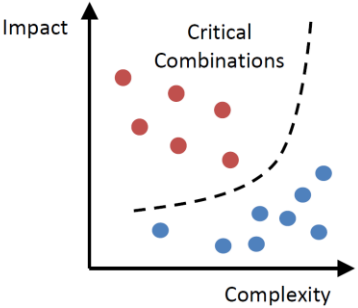
Main steps in risk management

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



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

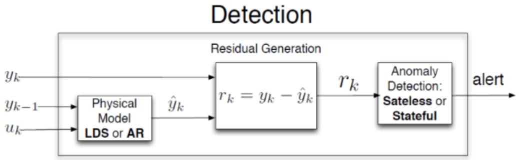
Tools for Risk Mitigation with Examples

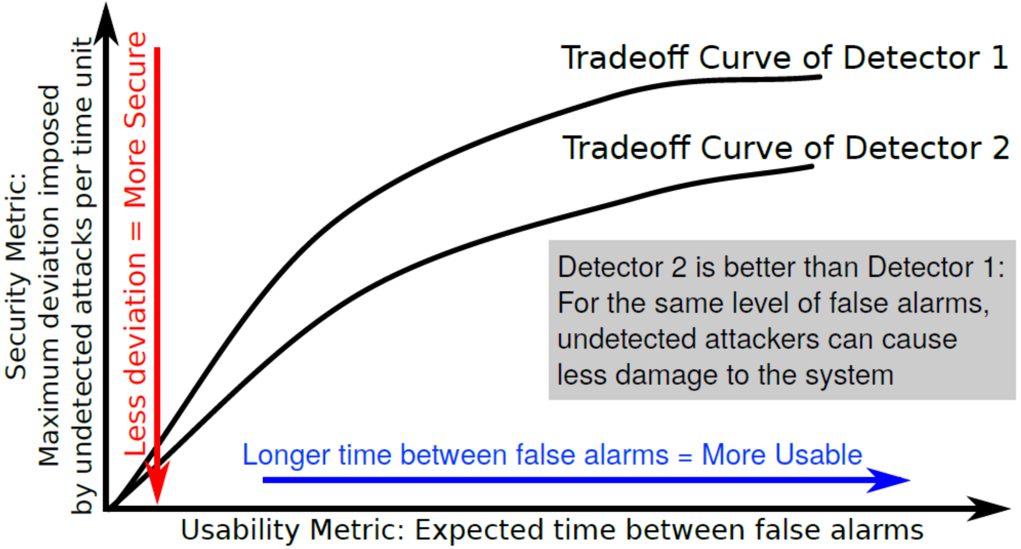


- **Prevention** (decrease likelihood by reducing vulnerability)
  - Watermarking and Moving Target Defense **(Example C)**
  - Coding and Encryption Strategies **(Examples B)**
  - Rational Security Allocation **(Example D)**
  - Privacy-preservation by Noise Injection **([1]: work by Le Ny and coauthors)**
- **Detection** (continuous anomaly monitoring)
  - Tuning of Detector Thresholds **(Example A**)
  - Secure State Estimation **([1]: work by Tabuada and coauthors)**
  - Watermarking and Moving Target Defense **([1]: work by Sinopoli and coauthors)**
  - Methods Related to Robust Statistics **([1]: work by Ishii and coauthors)**
- **Treatment** (compensate for or neutralize detected attack)
  - Secure State Estimation **([1]: work by Tabuada and coauthors)**
  - Countering DoS Attacks **([1]: work by Tesi and coauthors)**
  - Methods Related to Robust Statistics **([1]: work by Ishii and coauthors)**
  - Controller update **(Part 2)**

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

Example A: New Performance Metric for ICS Anomaly Detection



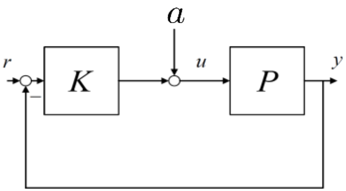


Detector 2 is better than Detector 1:
For the same level of false alarms,
undetected attackers can cause
less damage to the system

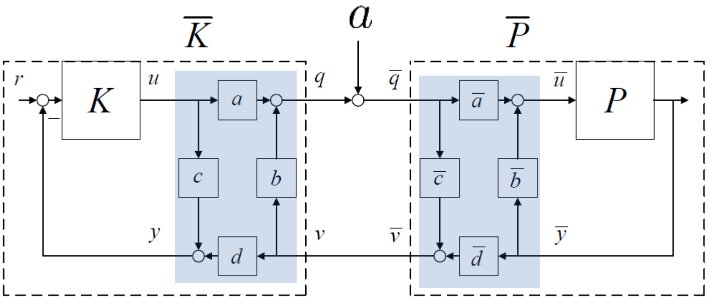
[Urbina *et al.*, CCS ‘16]

Example B: Two-Way Coding Counteracting Critical Undetectable Attacks

- Data injection attack a is undetectable if rate coincides with plant zeros $P(z_i) = 0$



- Introduce two-way coding (scattering transform) to render the zero dynamics harmless (z_i such that $\overline{P}(z_i) = 0$ are stable and very fast)



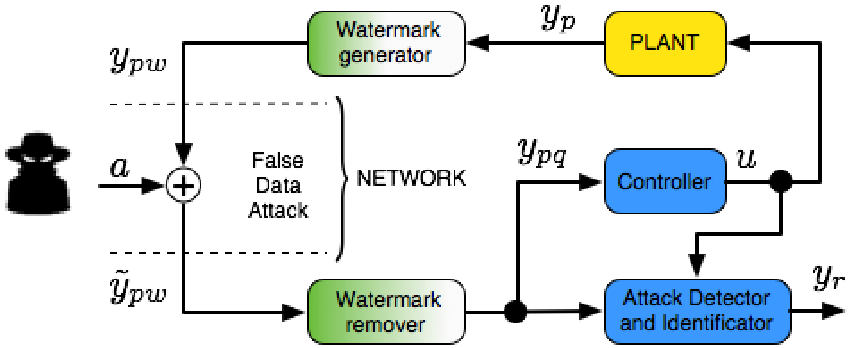
$$\begin{bmatrix} \overline{a} & \overline{d} \\ \overline{c} & \overline{d} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1}$$

- Similar result holds for sensor attacks and the poles of P

[Fang *et al.*, ICCPS, 2019]

Example C: Multiplicative Watermarking Counteracting Critical Undetectable Attacks

Watermark parameters change over time



- Multiplicative sensor watermarking can allow attack detection without degrading control performance

- Creates model asymmetry between attacker and operator

[Teixeira and Ferrari, ECC, 2018]

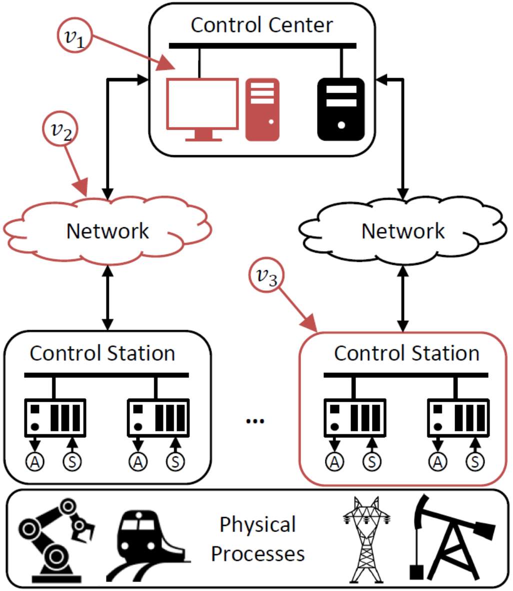
Example D: Rational Security Allocation

- Set of all security vulnerabilities

$$\mathcal{V} = \{v_1, v_2, \dots\}$$

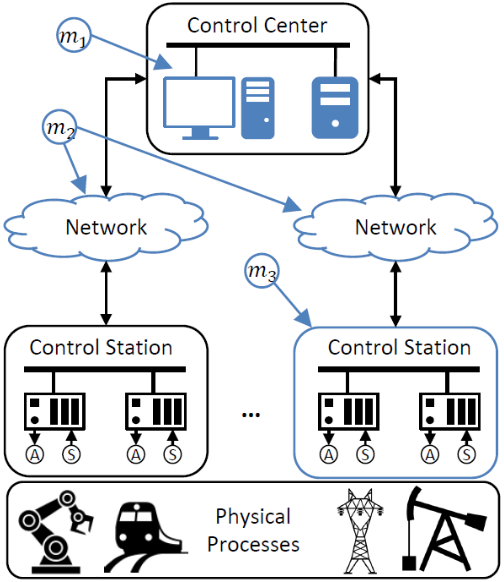
- $v \in \mathcal{V}$ can model (for instance)
    - Computers in control center do not have adequate protection (v_1)
    - Communication links are not encrypted or authenticated (v_2)

How to prevent high-risk attack scenarios involving exploitation of these vulnerabilities?



Security Measures

- Security measures $\mathcal{M} = \{m_1, m_2, \dots\}$
- $m \in \mathcal{M}$ can model (for instance)
  - Installing and maintaining anti-virus software in a computer
  - Encryption/authentication of a communication link
- Each $m \in \mathcal{M}$ assumed to prevent a subset of vulnerabilities $\mathcal{V}_m$
- The cost of implementing m is $c_m > 0$



Optimal Security Allocation Problem

- High risk attack scenarios $\mathcal{V}_a \subseteq \mathcal{V}$ have large *impact* and low *complexity*

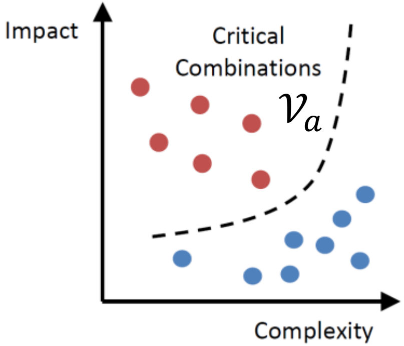
- Choose $\mathcal{M}_d^\star \subseteq \mathcal{M}$ solving

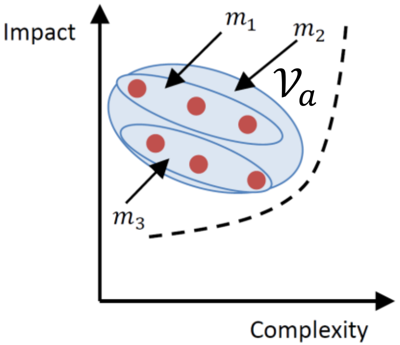
$$\min_{\mathcal{M}_d \subseteq \mathcal{M}} \sum_{m \in \mathcal{M}_d} c_m$$

such that all high-risk attacks scenarios $\mathcal{V}_a \subseteq \ \mathcal{V}$ prevented

- **Challenge 1:** Computing smallest $\mathcal{V}_a^\star \subseteq \mathcal{V}$ such that $\mathcal{M}_d$ prevents $\mathcal{V}_a^\star \Rightarrow \mathcal{M}_d$ prevents all $\mathcal{V}_a$

- **Challenge 2:** Solving for $\mathcal{M}_d^\star \subseteq \mathcal{M}$ (NP-hard)

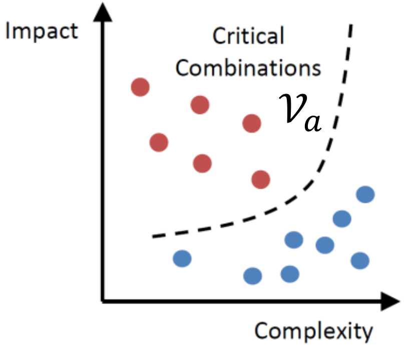




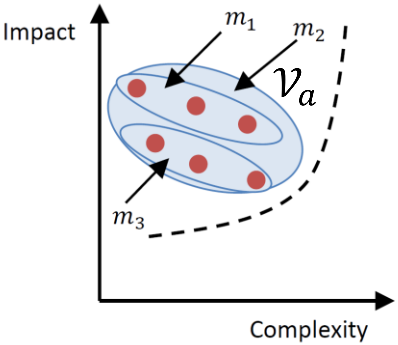
[Milošević *et al*., IJRNC, 2018]

Optimal Security Allocation Problem (cont'd)

- **Challenge 1:** Computing smallest $\mathcal{V}_a^\star \subseteq \mathcal{V}$ such that $\mathcal{M}_d$ prevents $\mathcal{V}_a^\star \Rightarrow \mathcal{M}_d$ prevents all $\mathcal{V}_a$
- **Approach:** Efficient pruning of search tree (in worst case, solution in exponential time)

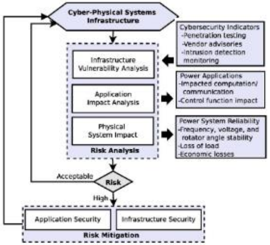


- **Challenge 2:** Solving for $\mathcal{M}_d^\star \subseteq \mathcal{M}$ (NP-hard)
- **Approach:** Exploit submodular structure to obtain approximate solution with optimality bound (solution in polynomial time)



[Milošević *et al.*, SafeThings, 2017, IJNRC 2018]

Summary and Outlook



- Cyber-secure control systems is an area of rapidly increasing importance
  - Most papers in tutorial paper less than 10 years old
  - IT security (still) necessary. Apply defense in depth!
- Careful and repeated risk analysis identifying the most relevant attacks is good starting point for secure control design – Tools for undetectable attacks used as recurring example in this presentation
- Careful attacker and operator modeling necessary – Many tools and solutions very sensitive to changes in the agents' resources
- **Topics for future work**
  - Tools from and for Machine Learning and AI
  - Fundamental design trade-offs (control performance – security – safety – privacy)
  - Further connections to fault-tolerant control and fault detection
  - Discrete-event systems