ROYAL INSTITUTE
OF TECHNOLOGY

Security and Limitations of Cyber-Physical Systems

Lecture 1: Introduction

Henrik Sandberg **André Teixeira**

Department of Automatic Control
ACCESS Linnaeus Centre, KTH Royal Institute of Technology
Stockholm, Sweden

Linköping University
August 24-26, 2015

ROYAL INSTITUTE OF TECHNOLOGY

Acknowledgments

- György Dán (KTH)
- Karl Henrik Johansson (KTH)

- Kin Cheong Sou (Chalmers)
- Iman Shames (Univ. Melbourne)

- Julien M. Hendrickx (UC Louvain)
- Raphael M. Jungers (UC Louvain)

ROYAL INSTITUTE OF TECHNOLOGY

Course Outline

- Monday (8:30-10:30):
  - Lecture 1 (HS): Introduction, data attacks against non-dynamical systems, power network monitoring, security index, graph min cut

- Tuesday (8:30-12:30):
  - Lecture 2 (HS): Attack space for cyber-physical systems: undetectable, stealth, covert, bias, DoS, replay attacks
  - Lecture 3 (AT): Defense mechanisms, risk management, anomaly detectors, watermarking

- Wednesday (8:30-10:30):
  - Exercise session: **Hand in exercises before to get credits**
  - Lecture 4 (HS): Physical limits of control implementations

ROYAL INSTITUTE OF TECHNOLOGY

Key References for Lecture 1

- André Teixeira, Kin Cheong Sou, Henrik Sandberg, Karl Henrik Johansson: "Secure Control Systems: A Quantitative Risk Management Approach". IEEE Control Systems Magazine, 35:1, pp. 24-45, February 2015.

- Julien M. Hendrickx, Karl Henrik Johansson, Raphael M. Jungers, Henrik Sandberg, Kin Cheong Sou: "Efficient Computations of a Security Index for False Data Attacks in Power Networks". IEEE Transactions on Automatic Control: Special Issue on Control of Cyber-Physical Systems, 59:12, pp. 3194-3208, December 2014.

ROYAL INSTITUTE OF TECHNOLOGY

Lecture 1

- Background and motivation
- Adversaries in networked control systems
- Quantifying security: A case study in power system monitoring
- The security index, and its computation

ICS-CERT MONITOR

January – April 2014

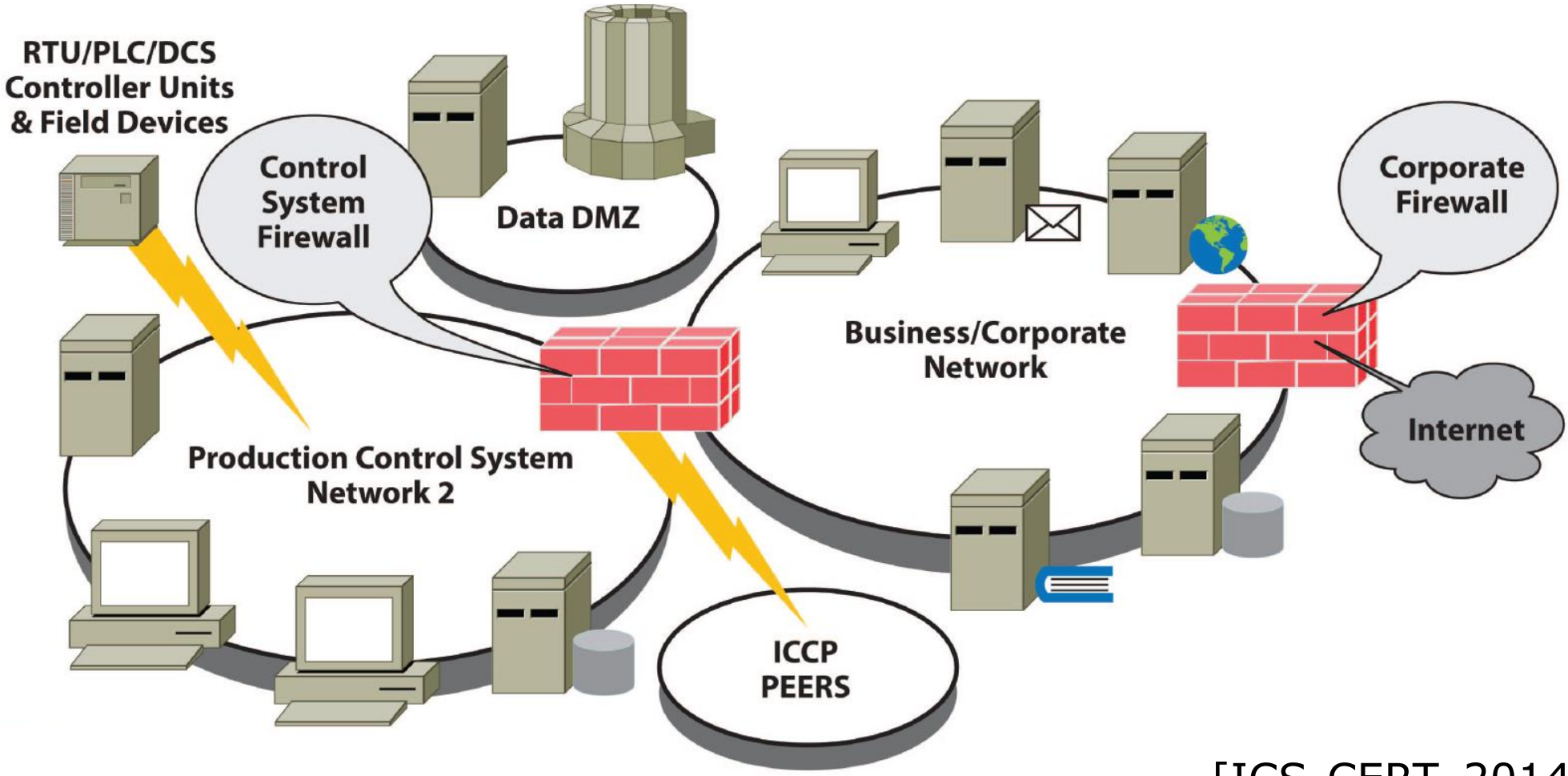
INCIDENT RESPONSE ACTIVITY

INTERNET ACCESSIBLE CONTROL SYSTEMS AT RISK

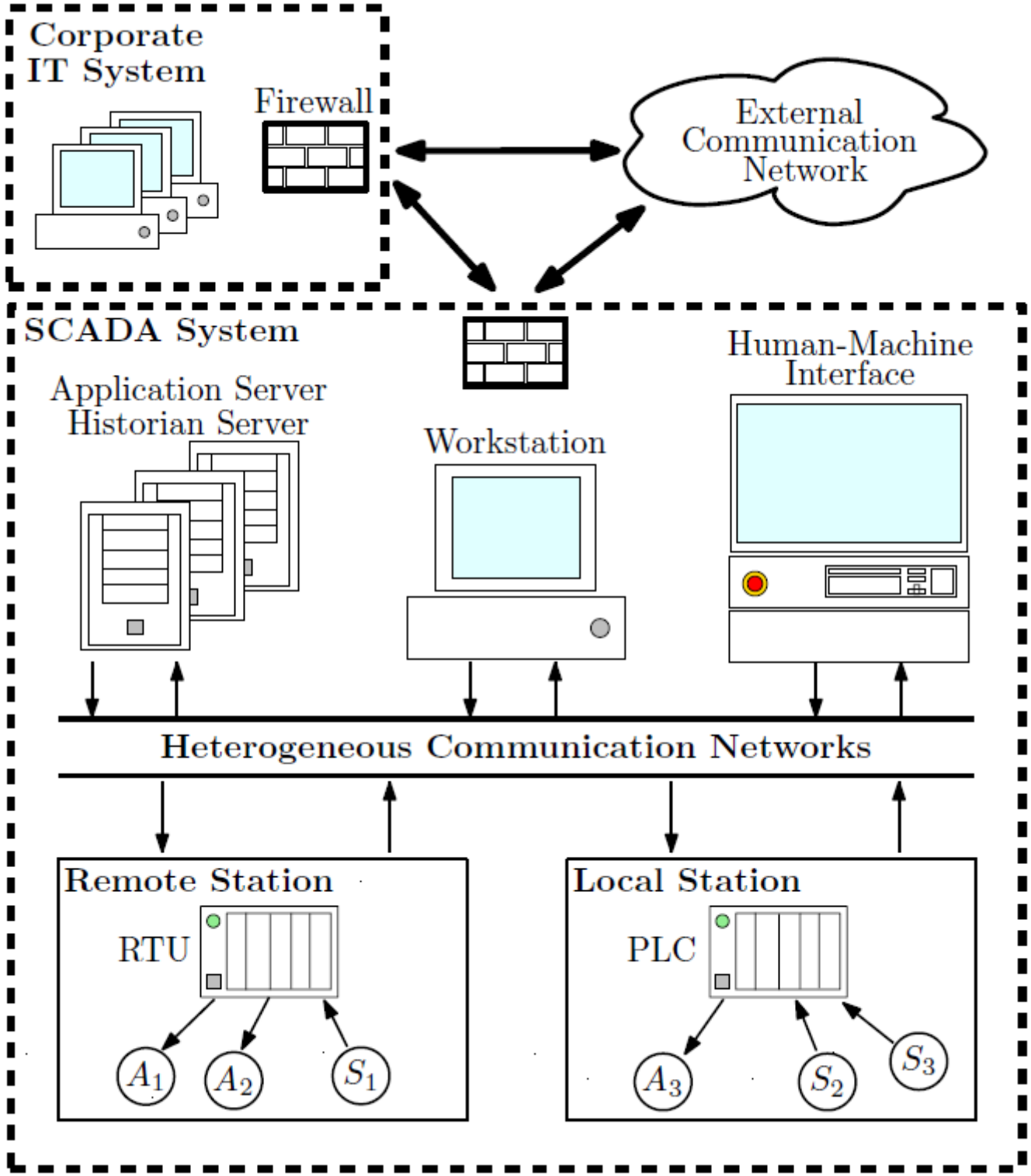
Is your control system accessible directly from the Internet? Do you use remote access features to log into your control system network? Are you unsure of the security measures that protect your remote access services? If your answer was yes to any or all these questions, you are at increased risk of cyber attacks including scanning, probes, brute force attempts and unauthorized access to your control environment.

ICS-CERT = Industrial Control Systems Cyber Emergency Response Team (https://ics-cert.us-cert.gov/)
Part of US Department of Homeland Security

Example 1: Industrial Control System (ICS) Infrastructure



[ICS-CERT, 2014]


Corporate
IT System
Firewall
External
Communication
Network
SCADA System
Application Server
Historian Server
Workstation
Human-Machine
Interface
Heterogeneous Communication Networks
Remote Station
RTU
A_1
A_2
S_1
Local Station
PLC
A_3
S_2
S_3

Example 2: The Stuxnet Worm (2010)

- **Targets:** Windows, ICS, and PLCs connected to variable-frequency drives
- Exploited **4 zero-day flaws**

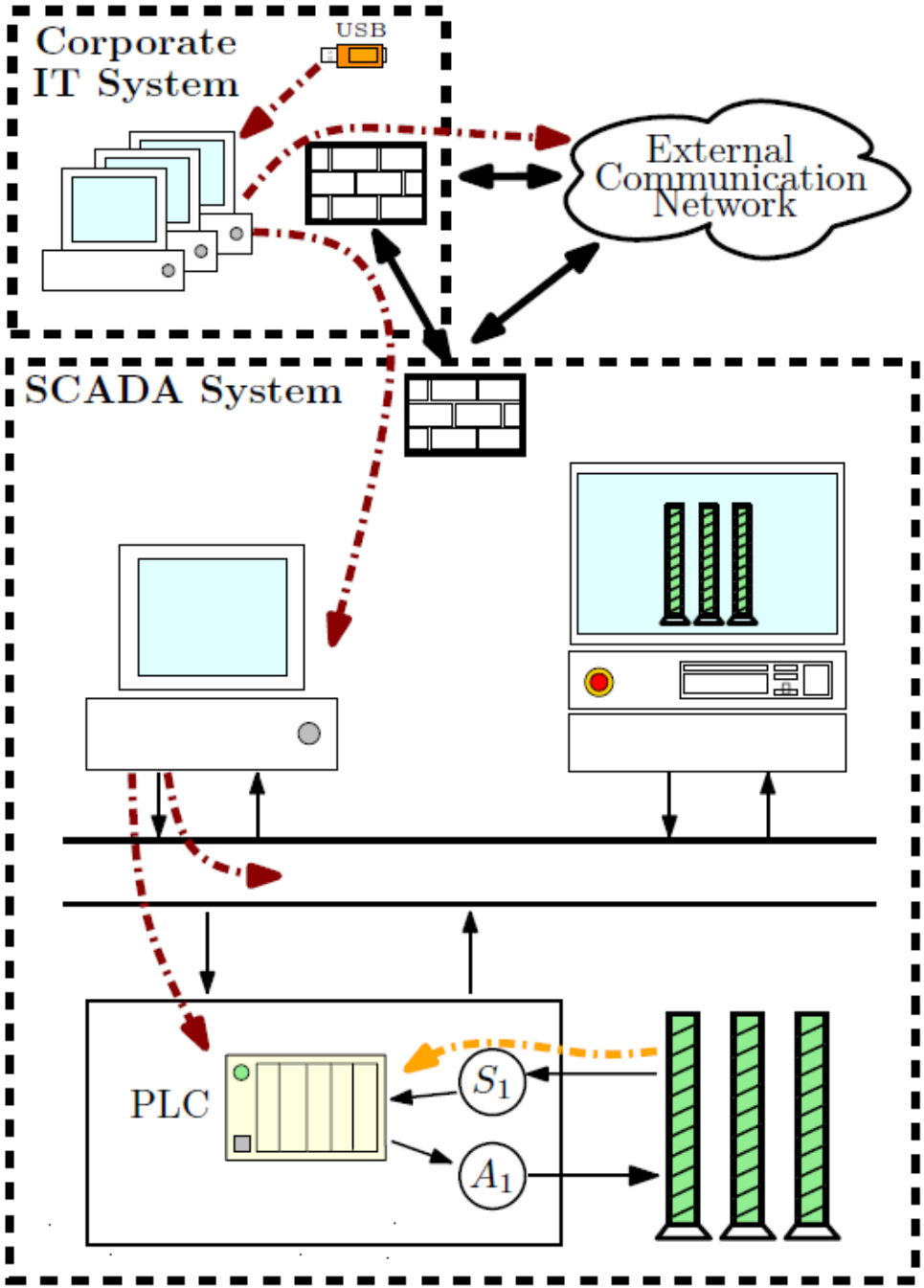
- **Speculated goal:**

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]



(a) Infection and data recording.

ROYAL INSTITUTE OF TECHNOLOGY

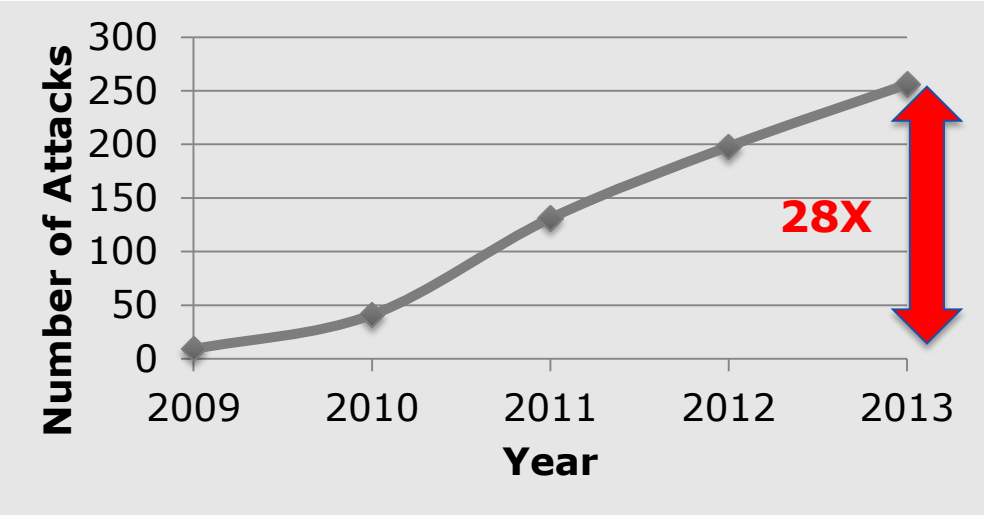
Example 3: Attack Demo from SPARKS

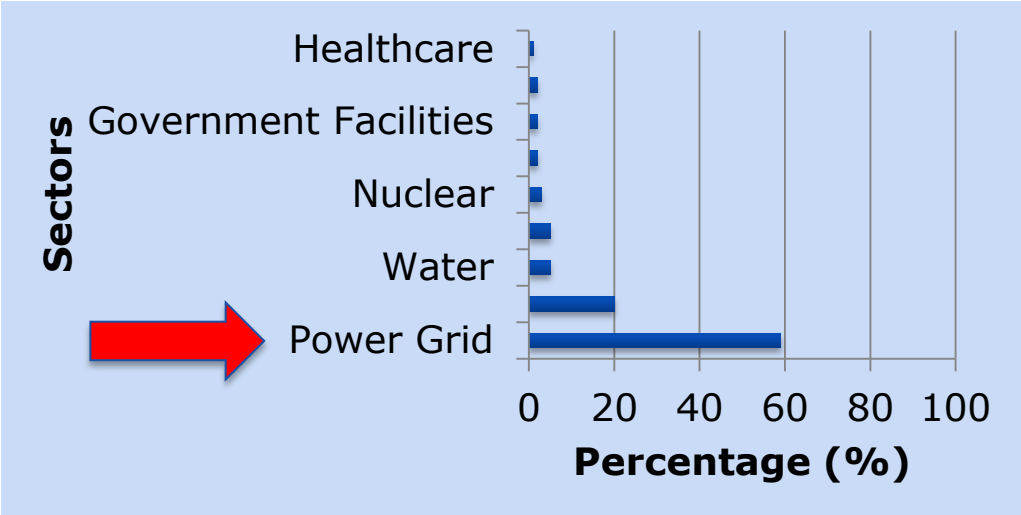
- SPARKS: EU FP7 project
  - https://project-sparks.eu/
- A Multi-stage Cyber-attack Demonstration
  - https://project-sparks.eu/publications/stakeholder-workshops/2nd-sparks-stakeholder-workshop/

At the stakeholder workshop, a multi-stage cyber-attack demonstration was presented to the stakeholders. The aim of this demonstration was to highlight the nature of the cybersecurity problem for the smart grid. The attack demonstration implemented a number of steps, such as a social engineering attack, and exploited known software vulnerabilities to realise a man-in-the-middle attack between a IEC 61850 client and a photovoltaic inverter located at the AIT smartEST lab in Vienna, Austria.

Some Statistics

Cyber incidents in critical infrastructures in the US
(Voluntarily reported to ICS-CERT)





[ICS-CERT, 2013]
[S. Zonouz, 2014]

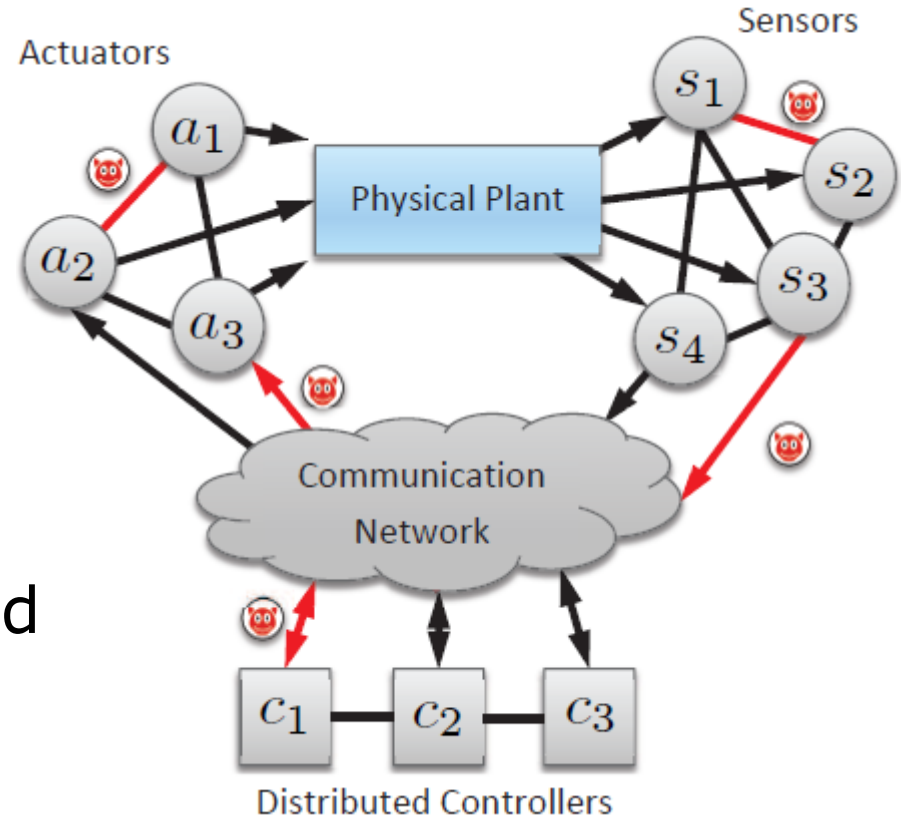
Cyber-Secure Control

Networked control systems

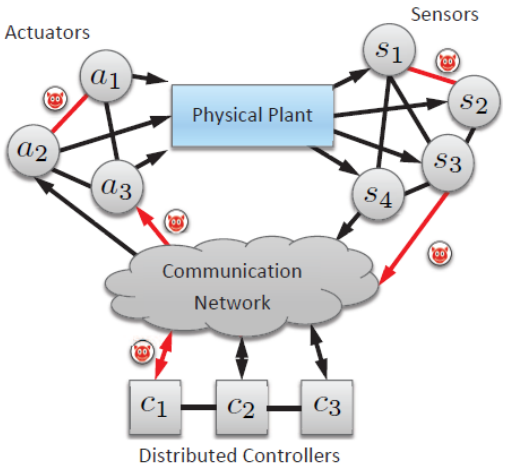
- are being **integrated with business/corporate networks**
- have many potential points of **cyber-physical attack**

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?



Special Controls Perspective Needed?



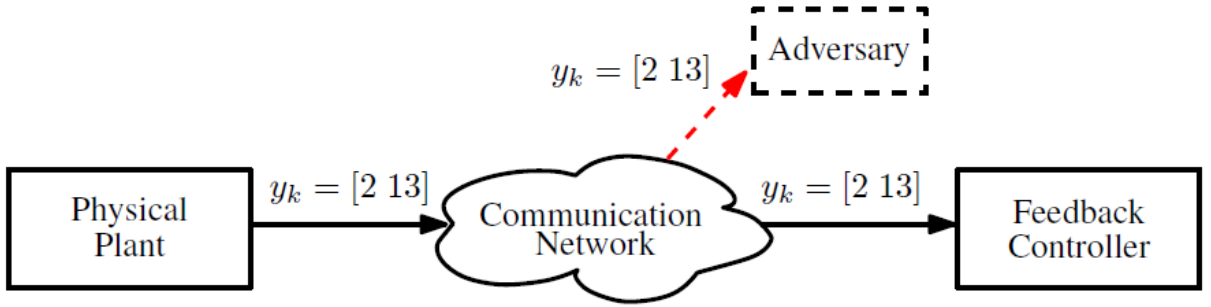
- Clearly IT security is needed: Authentication, encryption, firewalls, etc.

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
- Can we trust the interfaces and channels are really secured? (see OpenSSL Heartbleed bug…)

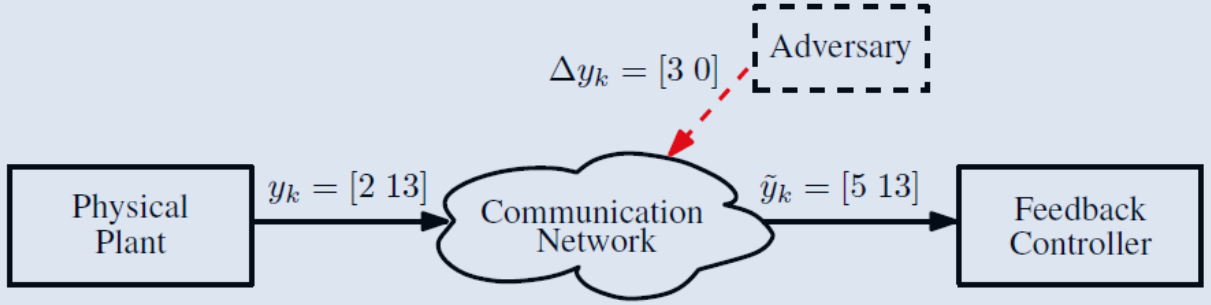
CIA in IT Security [Bishop, 2002]

- **C** – Confidentiality



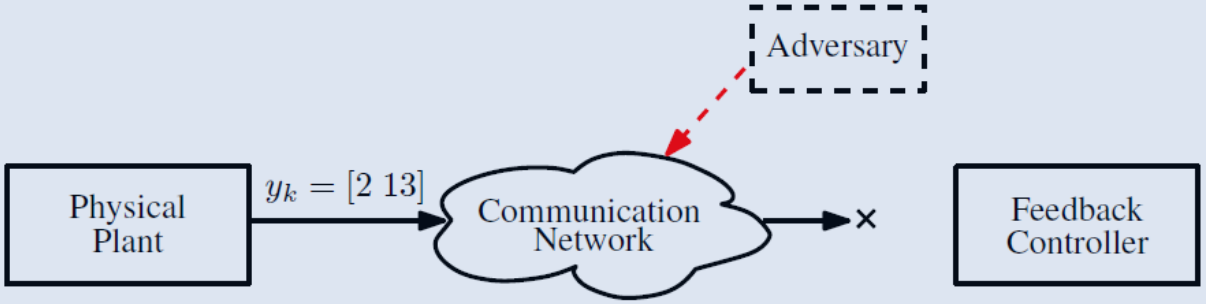
(a) Data confidentiality violation by a disclosure attack.

- **I** – Integrity



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

- **A** – Availability



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

ROYAL INSTITUTE OF TECHNOLOGY

Practical Challenges and Opportunities

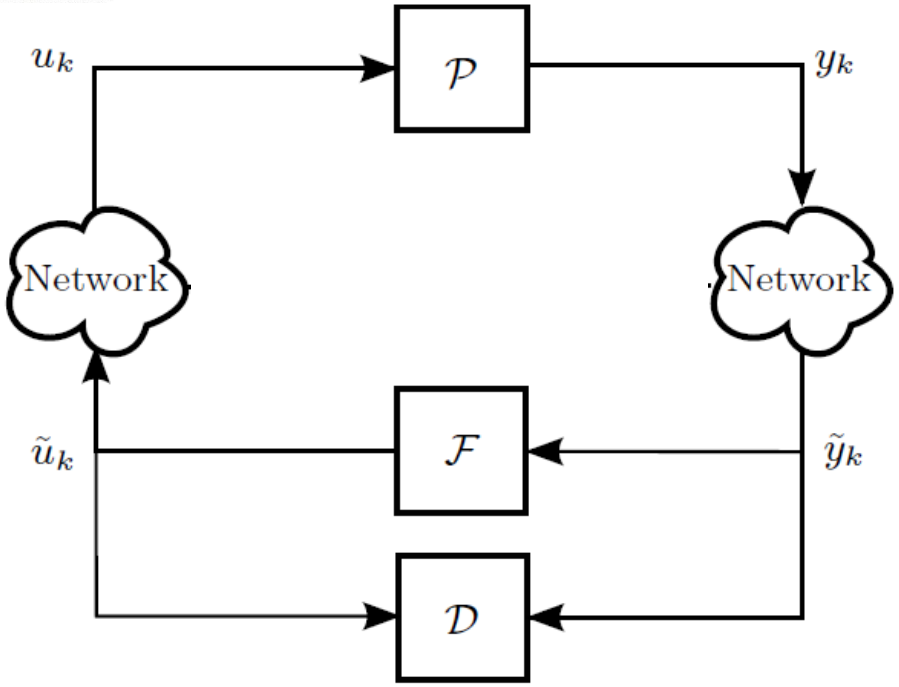
- Industrial Control Systems (ICSs) often a mix of old and new equipment. Improvements have to be **incremental**
- Many ICSs cannot easily be rebooted and patched with security updates
- How to best **collaborate** with **IT security** area? Strive for “defense in depth”: Run several independent **security systems in parallel**

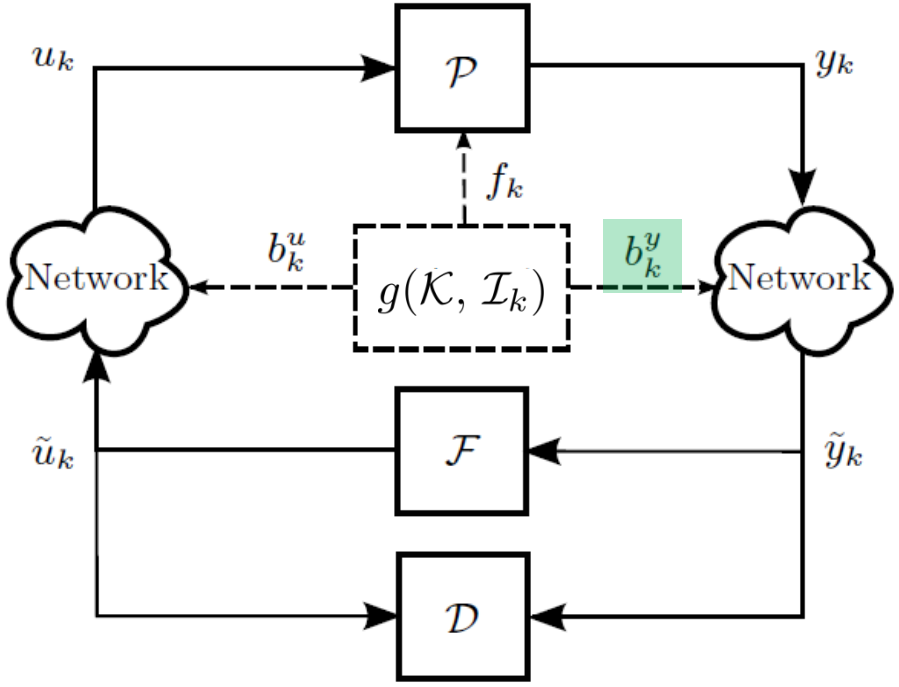
ROYAL INSTITUTE OF TECHNOLOGY

Lecture 1

- Background and motivation
- Adversaries in networked control systems
  - Which aspects are important to consider?
- Quantifying security: A case study in power system monitoring
- The security index, and its computation

Networked Control System under Attack





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

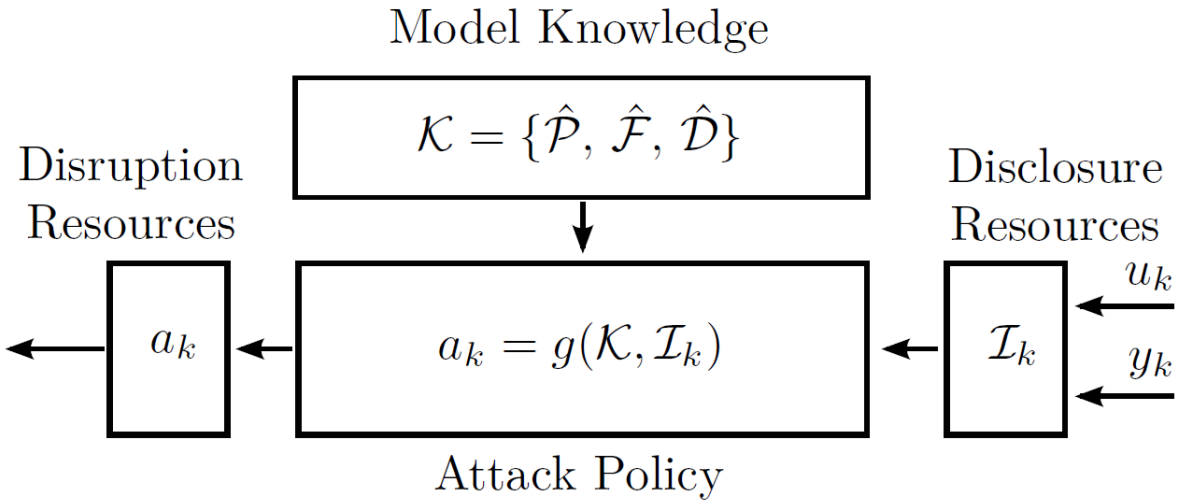
- Physical Attacks f_k
- Deception Attacks

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

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

Case study later today

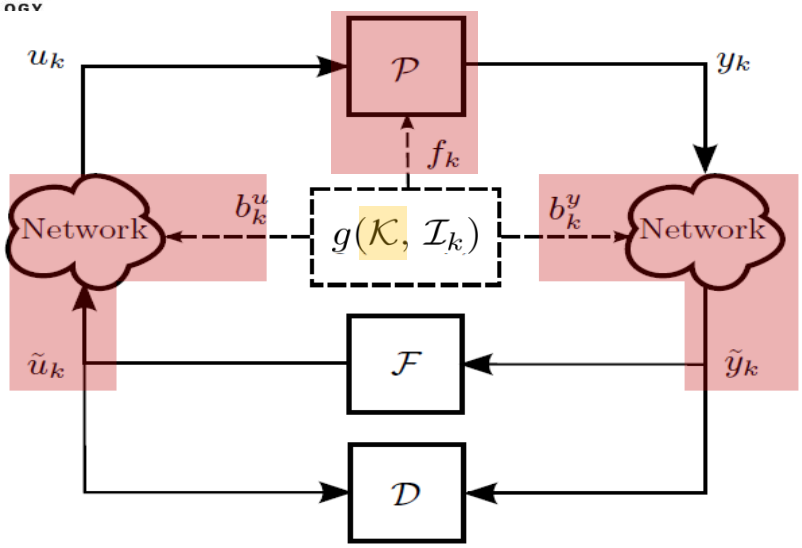
Adversary Model

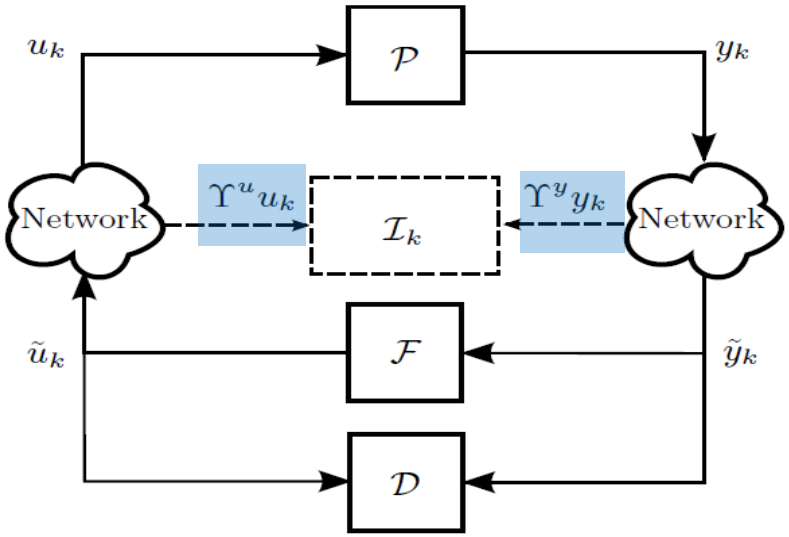


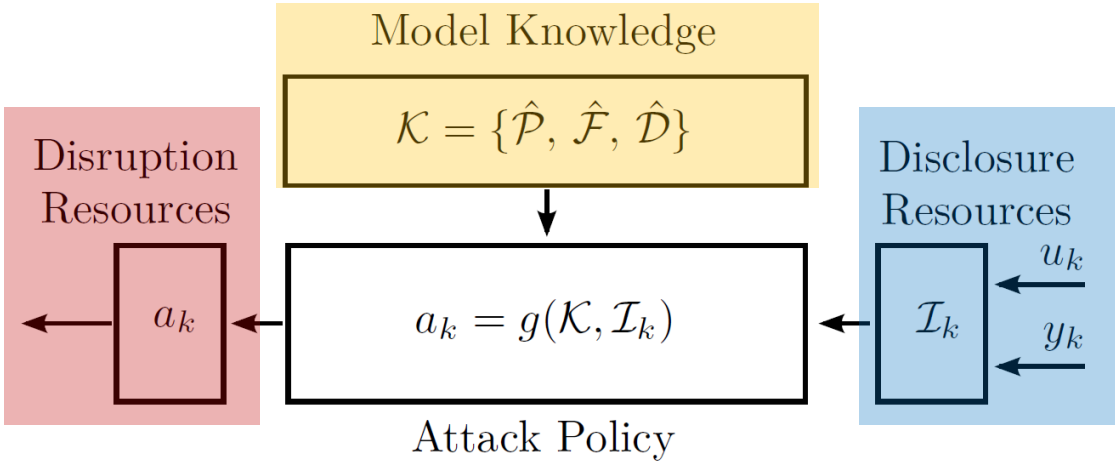
- **Attack policy:** Goal of the attack? Destroy equipment, increase costs,…
- **Model knowledge:** Adversary knows models of plant and controller? Possibility for stealthy attacks…
- **Disruption/disclosure resources:** Which channels can the adversary access?

[Teixeira *et al*., HiCoNS, 2012]

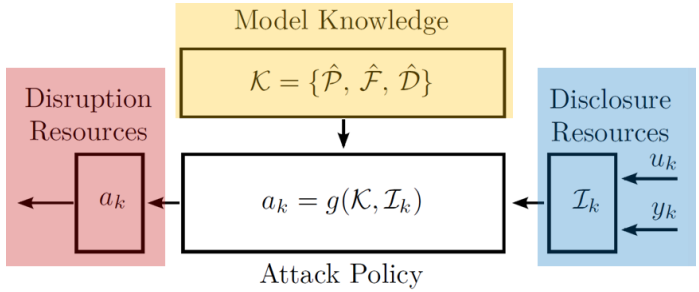
Networked Control System with Adversary Model

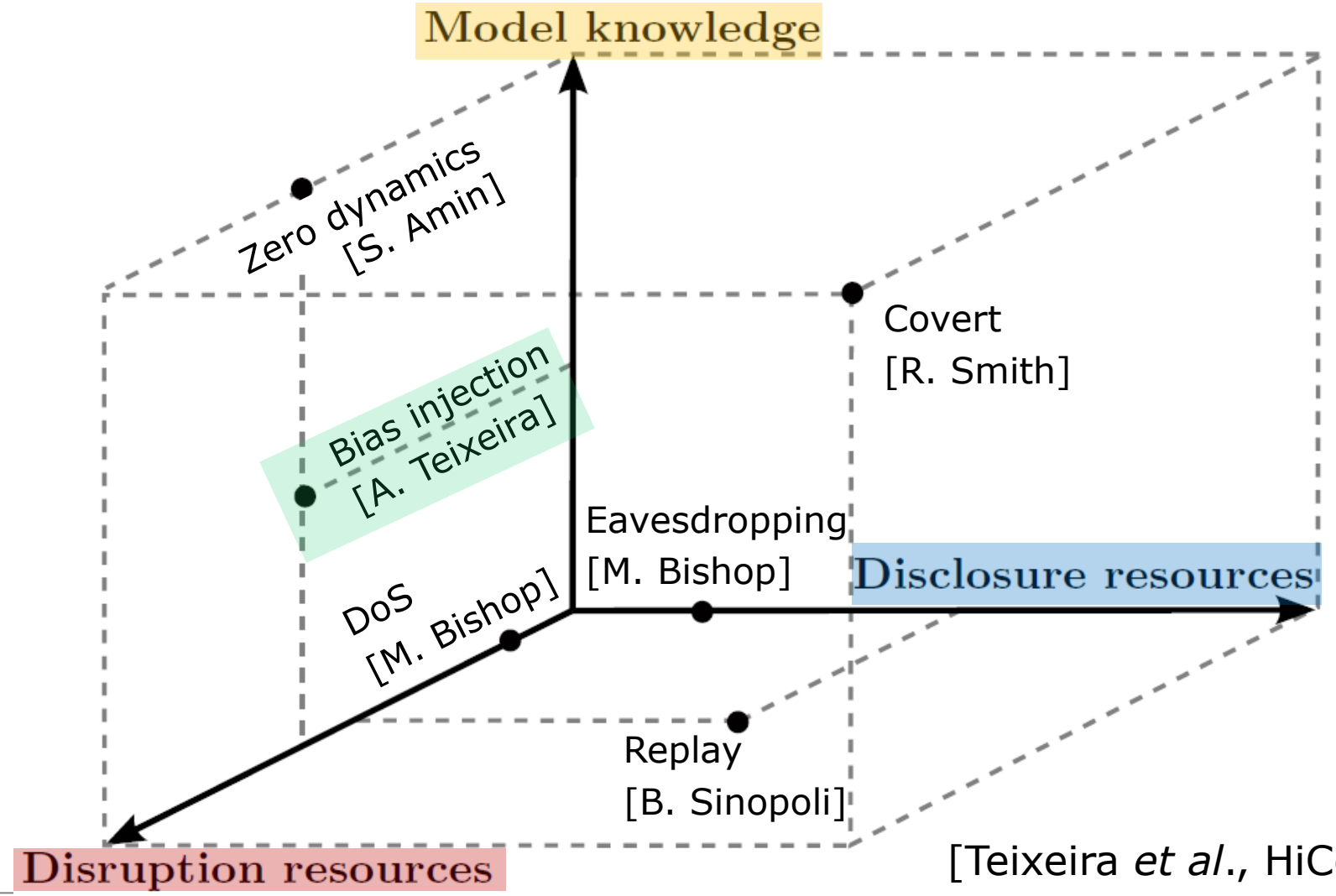






Attack Space





[Teixeira *et al.*, HiCoNS, 2012]

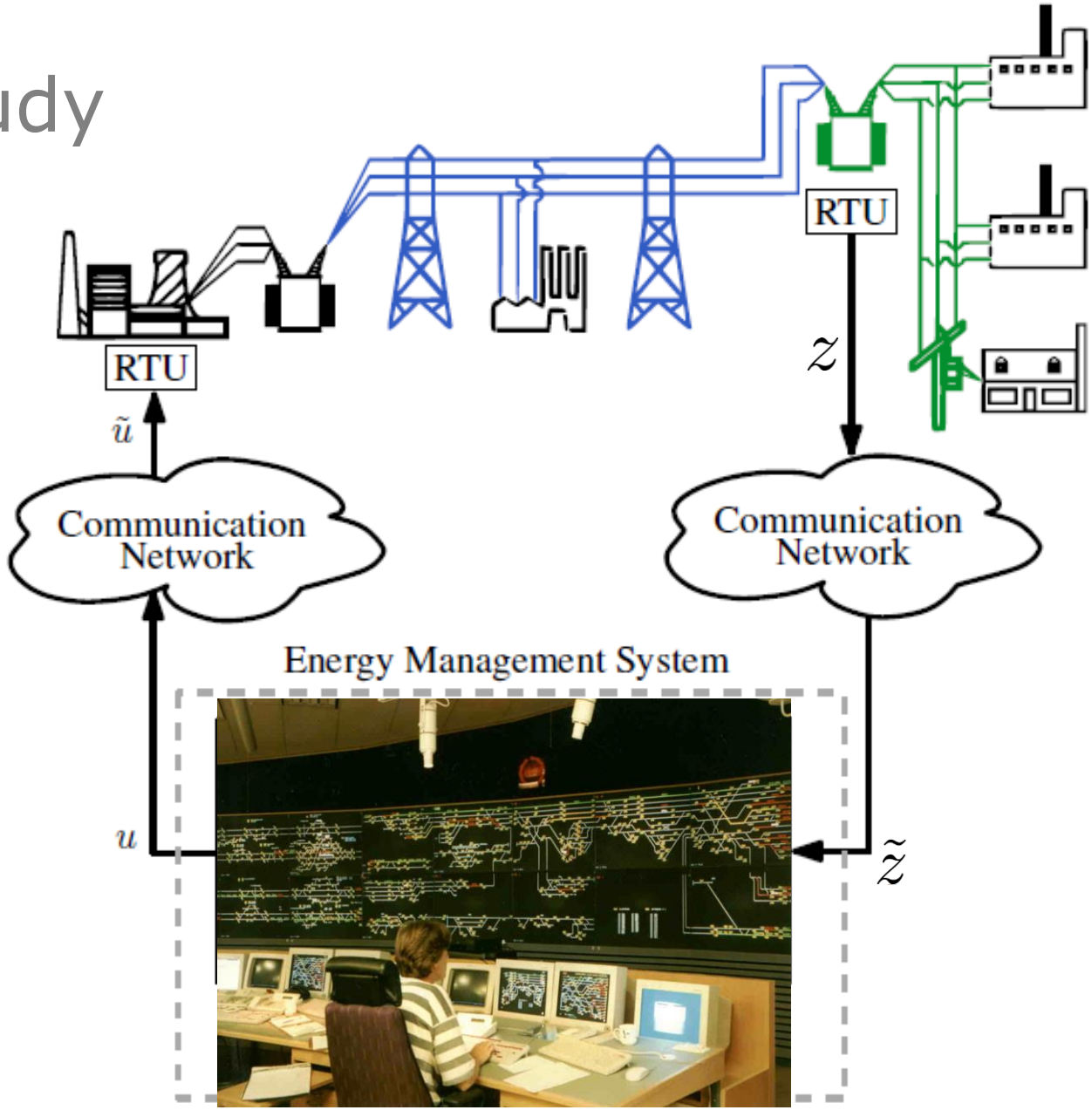
ROYAL INSTITUTE OF TECHNOLOGY

Lecture 1

- Background and motivation
- Adversaries in networked control systems
- Quantifying security: A case study in power system monitoring
- The security index, and its computation

Case Study

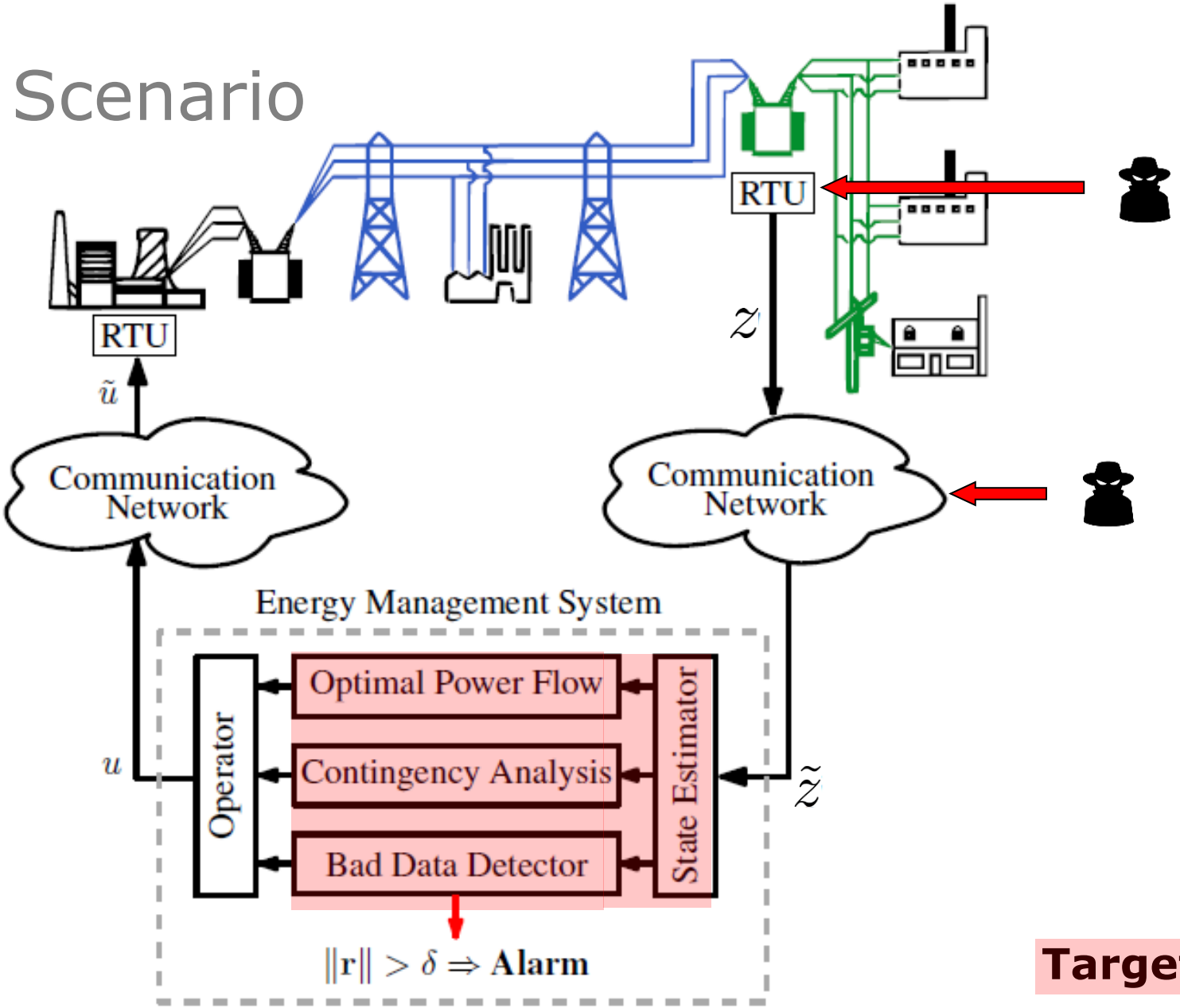
Transmission power system control and monitoring over networks



The Scenario

Practically motivated problem…

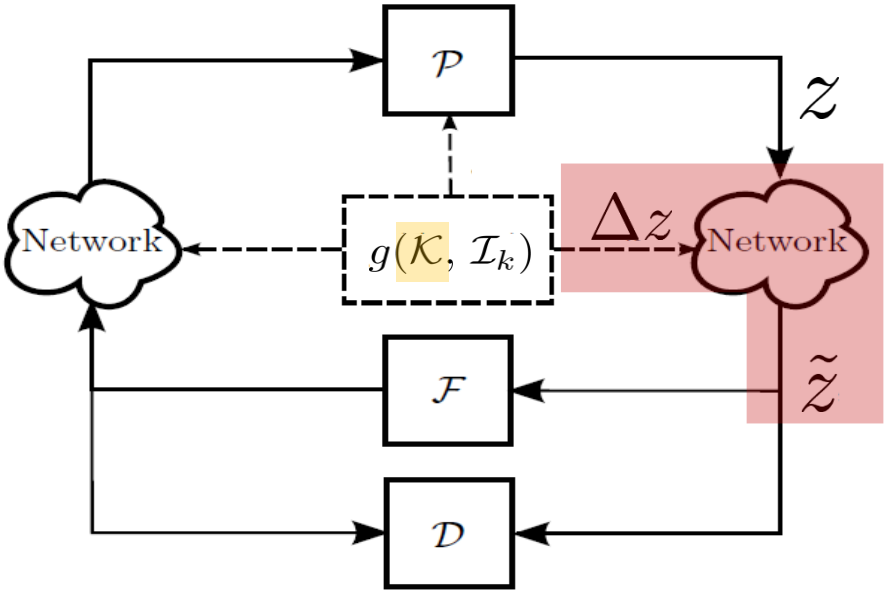
How much security does the Bad Data Detector provide?

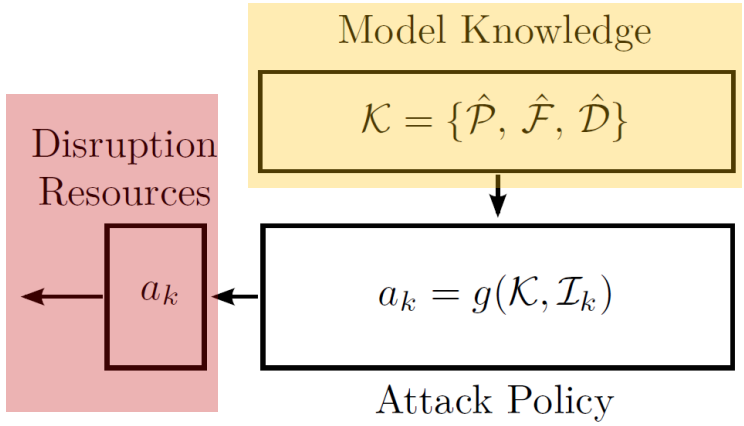


Target

[Giani *et al.*, IEEE ISRCS, 2009]
[Mohajerin Esfahani *et al.*, CDC, 2010]

Adversary Model





- **Attack policy:** Induce bias in power measurements without alarms
- **Model knowledge:** Steady-state model of power system
- **Disruption resources:** Small number of measurement channels

Can we quantify how hard such attacks would be?

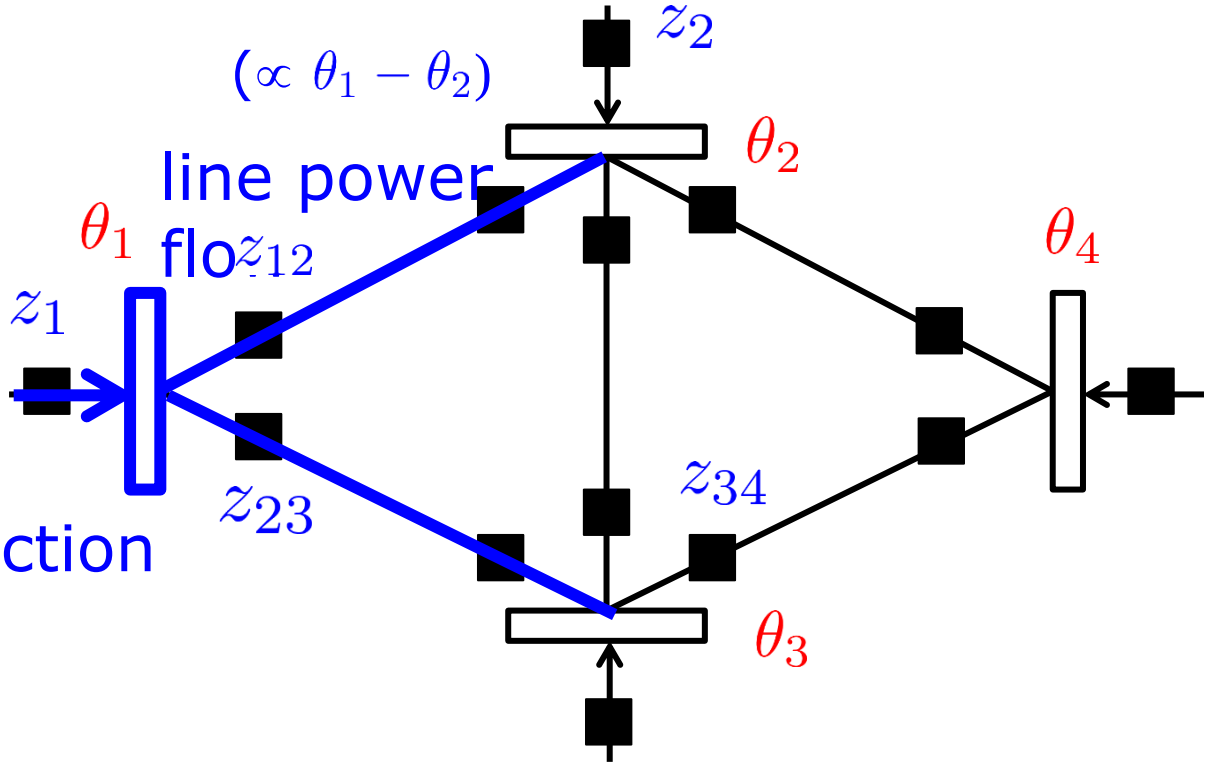
Steady-State Power System Model

States (θ)
= bus voltage phase angles

(flow conservation)

bus injection

Measurements (z)
= line power flow & bus injection



"DC power flow model":

$$z = H\theta$$

measurement matrix

bus (node)

line (edge)

meter

[Hendrickx *et al.*, IEEE TAC, 2014]

Structure of Measurement Matrix *H*

$$H = \begin{bmatrix} P_1 D A^T \\ -P_2 D A^T \\ P_3 A D A^T \end{bmatrix}$$

(flow measurements)
(flow measurements)
(injection measurements)

- A - directed incidence matrix of graph corresponding to power network topology
- D - nonsingular diagonal matrix containing reciprocals of reactance of transmission lines
- $P_.$ - measurement selection matrices (rows of identity matrices)
- More measurements than states. Redundancy!

[Hendrickx *et al*., IEEE TAC, 2014]

ROYAL INSTITUTE OF TECHNOLOGY

Example on the Board

State Estimation by Weighted Least Squares

State estimator (LS)

$$\tilde{z} = H_{2:}\theta_{2:} + \Delta z$$

$$\Rightarrow \hat{\theta}_{2:} = (H_{2:}^T W H_{2:})^{-1} H_{2:}^T W \tilde{z}$$

wrong

wrong

Contingency analysis

wrong

OPF calculations

What if the measurements were **wrong**?

$\tilde{z} = z + \Delta z$ → random measurement noise

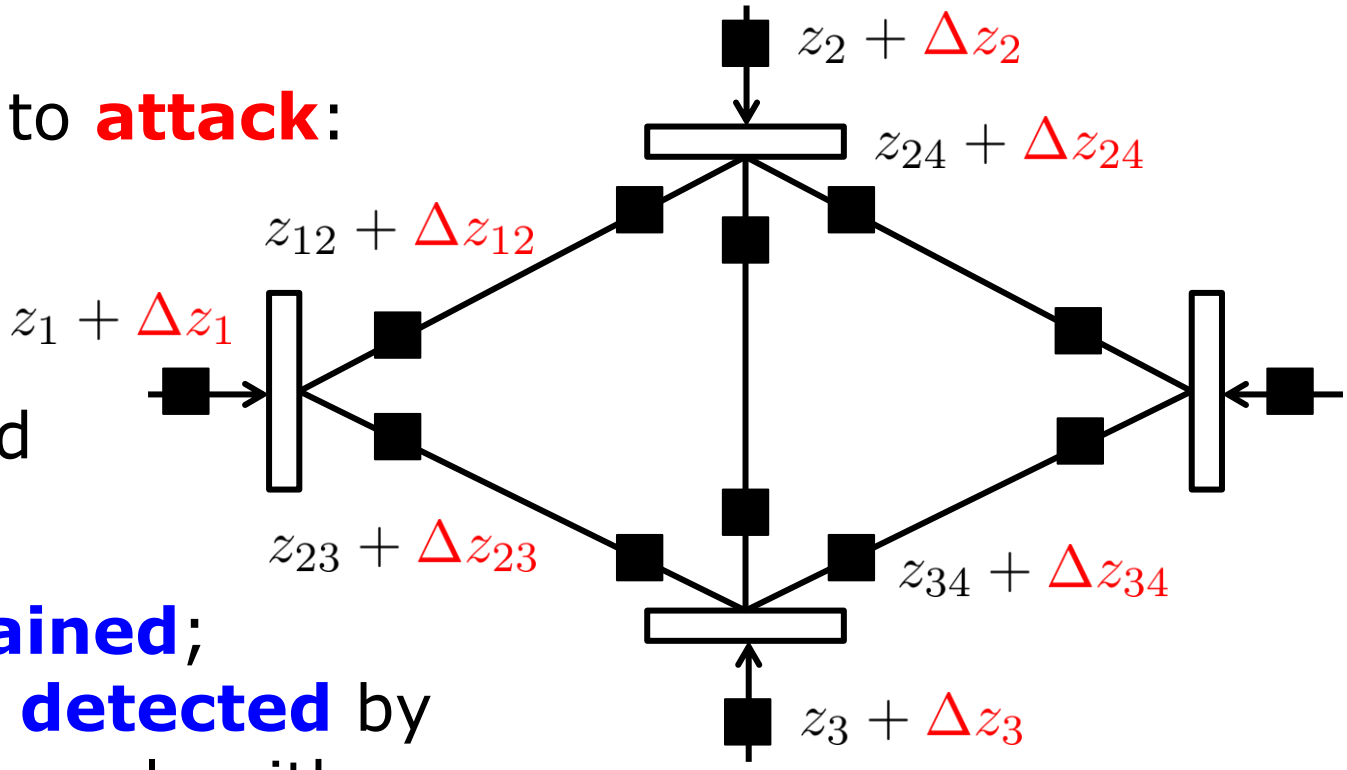
intentional data attack → $\hat{\theta}_{2:} = \theta_{2:} + \Delta\theta_{2:}$

Stealthy Additive Deception Attack

Measurements subject to **attack**:

$$\tilde{z} = z + \Delta z$$

Is there a state explaining the received measurements?

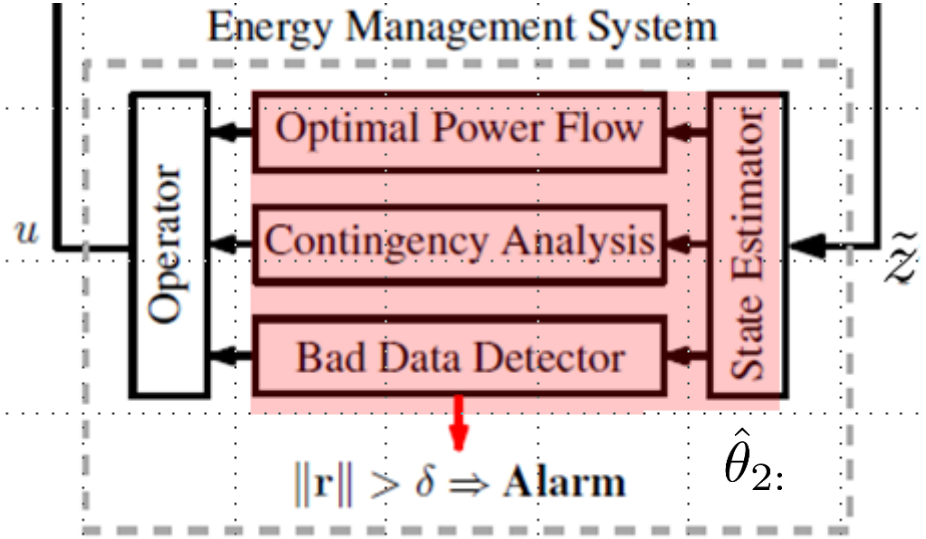


Attack is **constrained**; otherwise will be **detected** by Bad Data Detection algorithm

Stealth attack: $\Delta z = H\Delta\theta$, arbitrary $\Delta\theta \in \mathbb{R}^{n+1}$

[Liu *et al*., ACM CCCS, 2009], [Sandberg *et al*., CPSWEEK, 2010]

Bad Data Detection Algorithm



- $\mathbf{r} := \tilde{z} - H_{2:}\hat{\theta}_{2:}$

$$= \left(I - H_{2:}(H_{2:}^T W H_{2:})^{-1} H_{2:}^T W\right) \Delta z$$

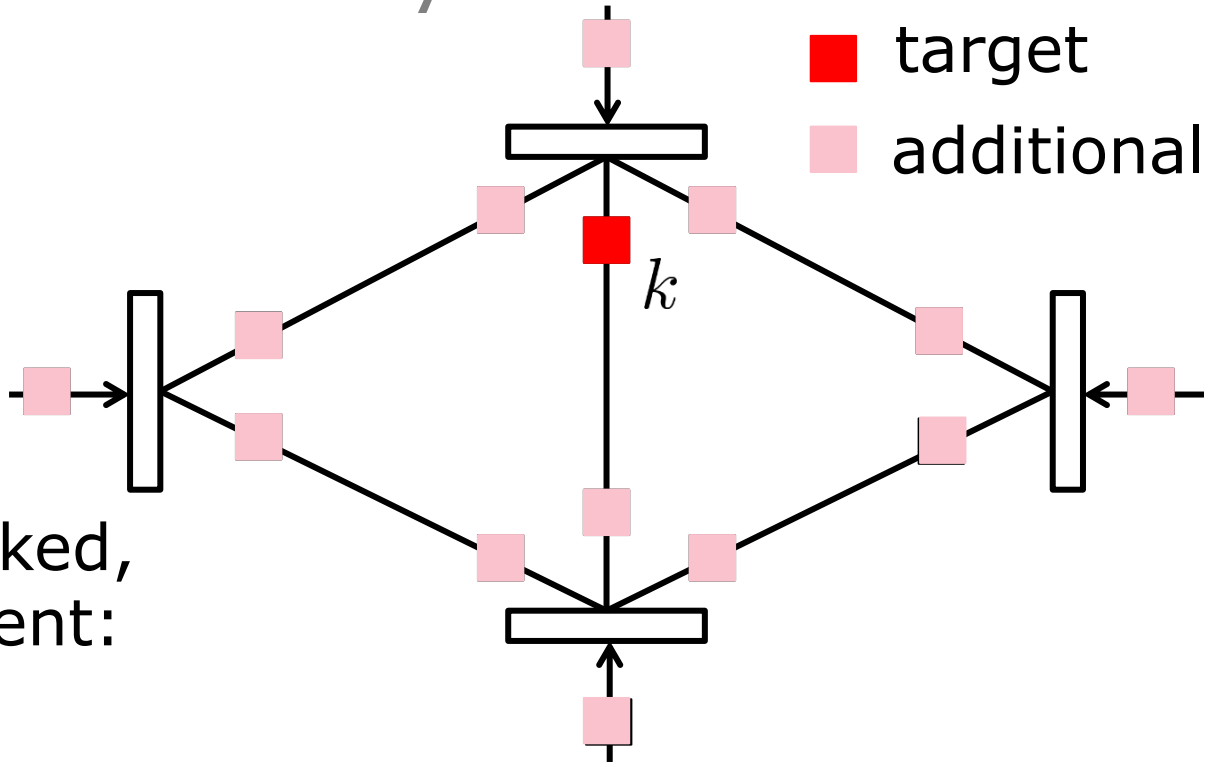
- Stealth attack: $\Delta z = H\Delta\theta$, arbitrary $\Delta\theta \in \mathbb{R}^{n+1}$

- $\mathbf{r} \equiv 0$, since $H\Delta\theta = H_{2:}\Delta\theta_{2:}$ for some $\Delta\theta_{2:} \in \mathbb{R}^n$

Quantification: Security Index

Stealth attack $\Delta z = H\Delta\theta$

In general, $e_k \notin \mathrm{span}(H)$



Minimum # of meters attacked, targeting the k^{th} measurement:

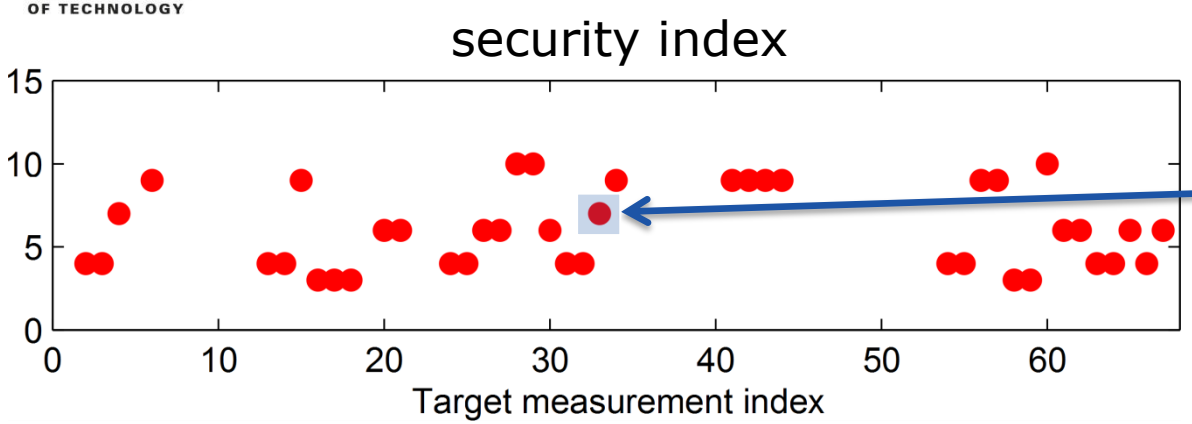
$$\min_{\Delta\theta\in\mathbb{R}^{n+1}} \ \mathrm{card}(H\Delta\theta)$$

$$\text{s.t. } H(k,:)\Delta\theta = 1$$

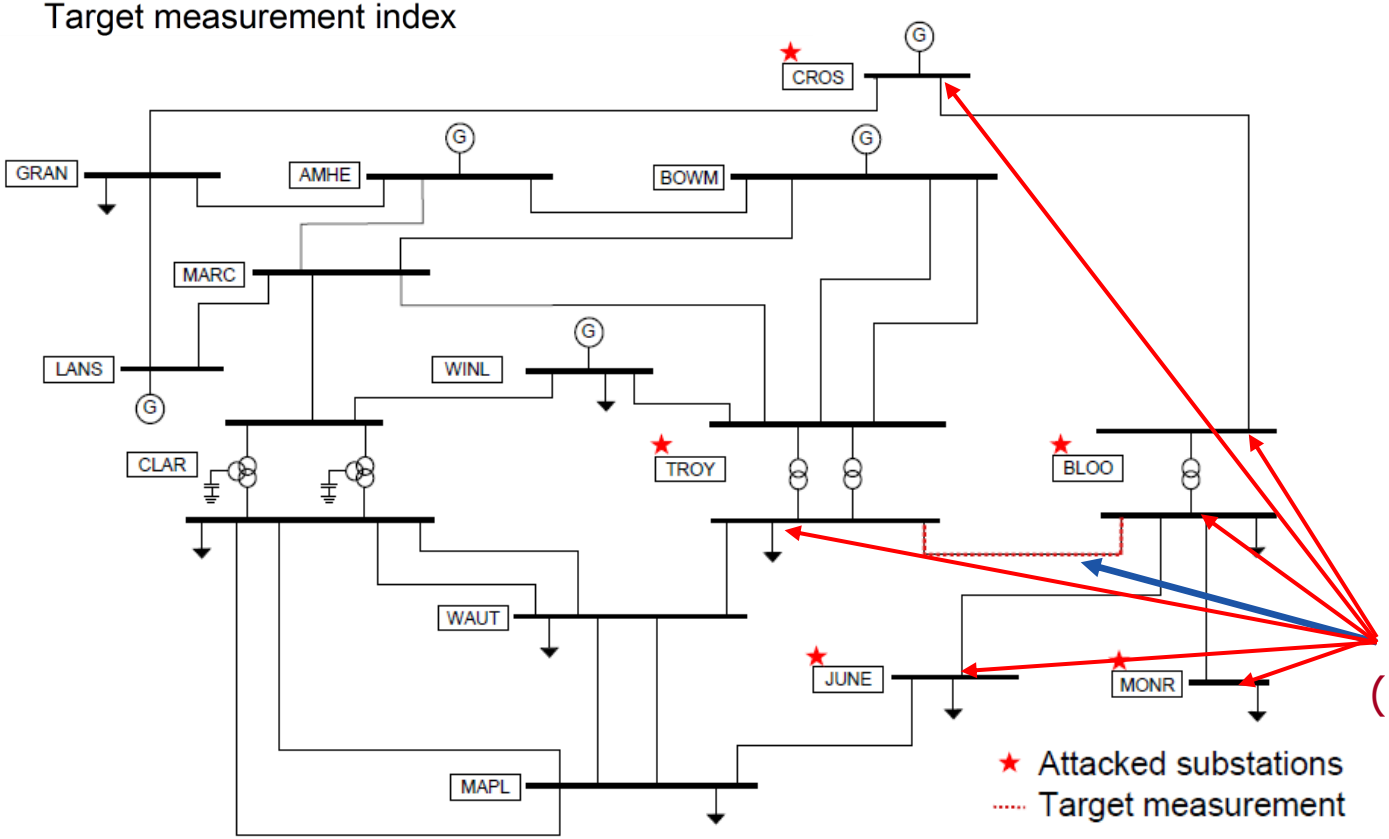
Minimum objective value = **security index**
[Sandberg *et al.*, CPSWEEK, 2010]

See also [Kosut *et al.*, IEEE TSG, 2011]

Security Indices for 40-bus Network

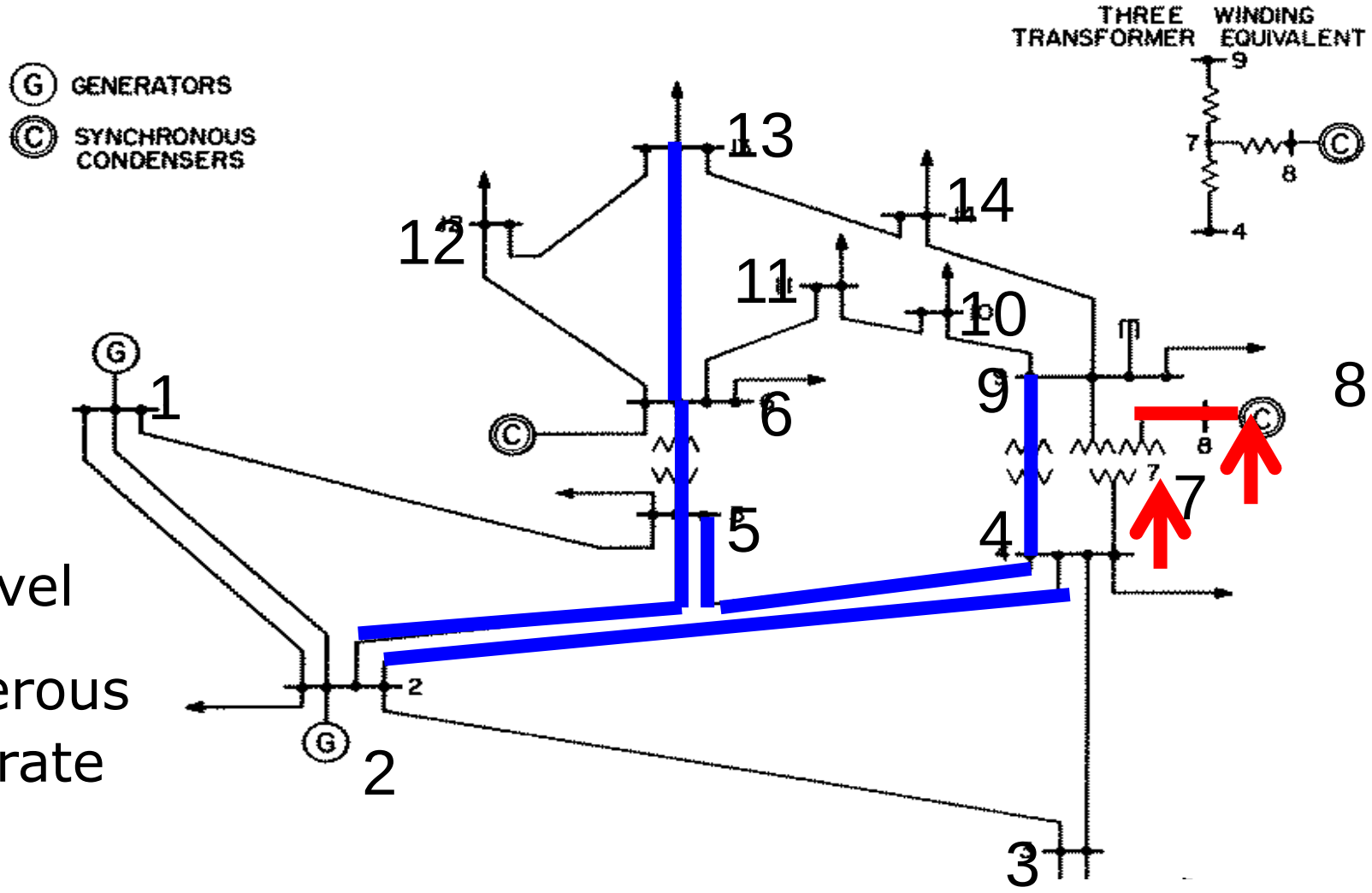


At least 7 measurements involved in a stealth attack against measurement 33



Attack 33
(7 measurements)

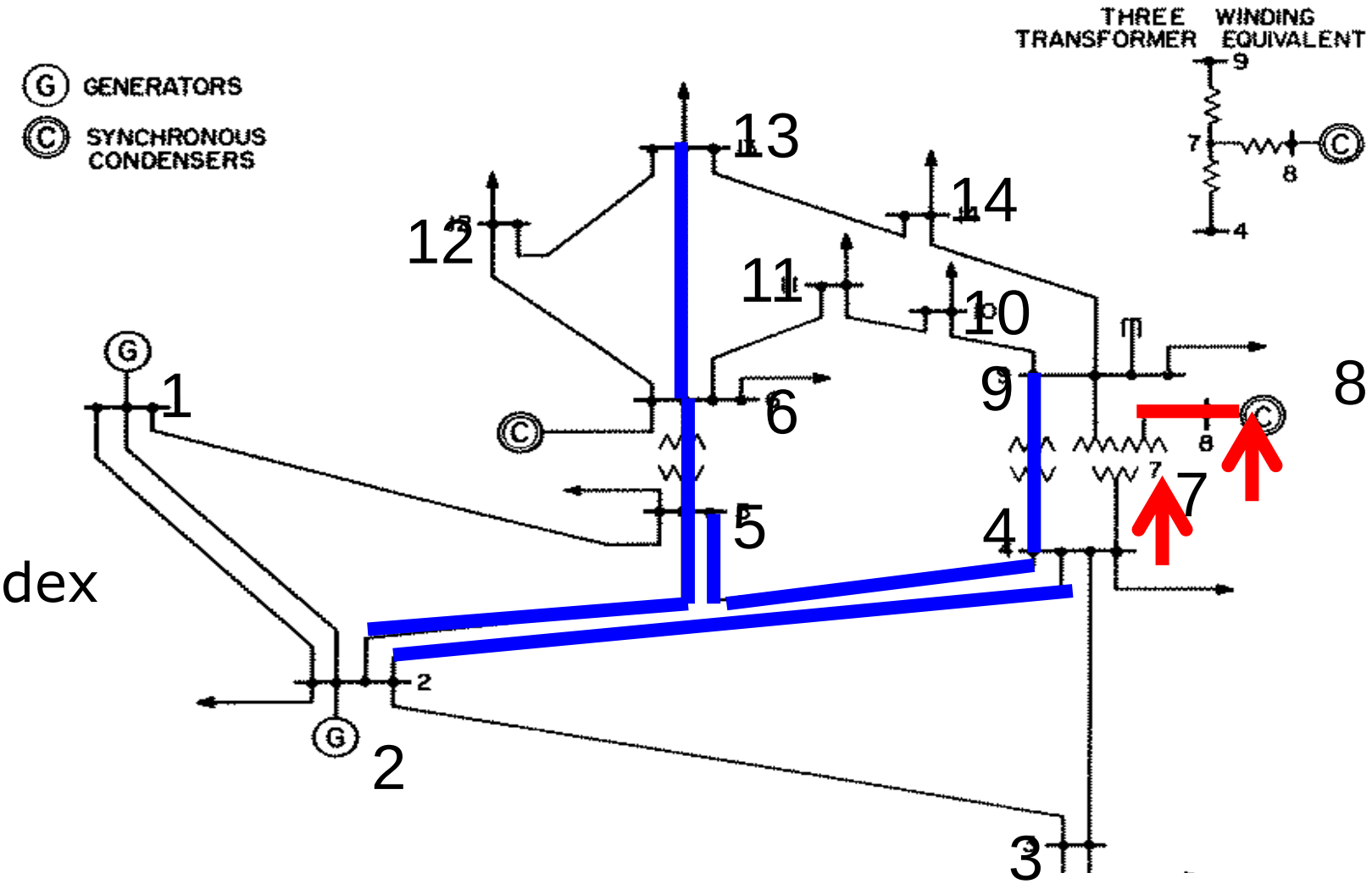
The Goal: Quantify Security to Aid Allocation of Protection



Security level

- dangerous
- moderate
- safe

IEEE 14 Bus Vulnerable Measurements



security index

- 4
- 7
- ≥ 10

ROYAL INSTITUTE OF TECHNOLOGY

Security index problem

$$\min_{\Delta\theta \in \mathbb{R}^{n+1}} \operatorname{card}(H\Delta\theta)$$

$$\text{s.t. } H(k,:)\Delta\theta = 1$$

How to solve?

Closely related to compressed sensing and computation of **cospark** of *H* [Tillmann and Pfetsch, IEEE TIT, 2013]. Problem known to be **NP-hard** for arbitrary *H*.

[Theorem 1, Hendrickx *et al*., IEEE TAC, 2014]

Security Index Computation – MILP

$$\min_{\Delta\theta \in \mathbb{R}^{n+1}} \ \mathrm{card}(H\Delta\theta)$$

$$\text{s.t.} \ \ H(k,:)\Delta\theta = 1$$

- Cardinality minimization problem
- Mixed integer linear program (MILP)
- Big *M* reformulation
- **Exact** solution (solver: CPLEX)
- Solution algorithm **not scalable**

$$\min_{\Delta\theta,\, z} \sum_i z(i)$$

$$\text{s.t.}$$

$$-Mz \le H\Delta\theta \le Mz$$

$$H(k,:)\Delta\theta = 1$$

$$z(i) \in \{0,1\} \quad \forall i$$

MILP formulation

(Exercise 1)

Security Index Computation – LASSO

$$\min_{\Delta\theta} \|H\Delta\theta\|_1$$

$$\text{s.t. } H(k,:)\Delta\theta = 1$$

- Convex linear program (LP)
- Known as LASSO
- **Approximate** solution
- Less expensive to solve

$$\min_{\Delta\theta,\, z} \sum_i z(i)$$

$$\text{s.t.}$$

$$-z \leq H\Delta\theta \leq z$$

$$H(k,:)\Delta\theta = 1$$

$$z(i) \in \mathbb{R} \quad \forall i$$

LP formulation

ROYAL INSTITUTE OF TECHNOLOGY

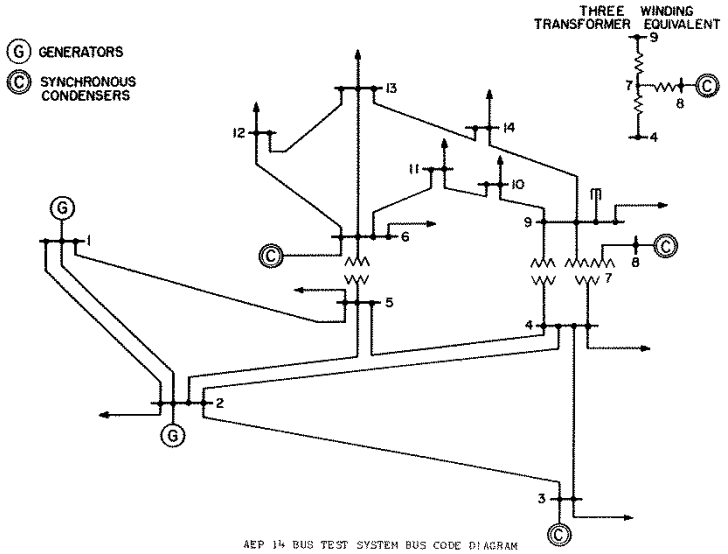
Wish List

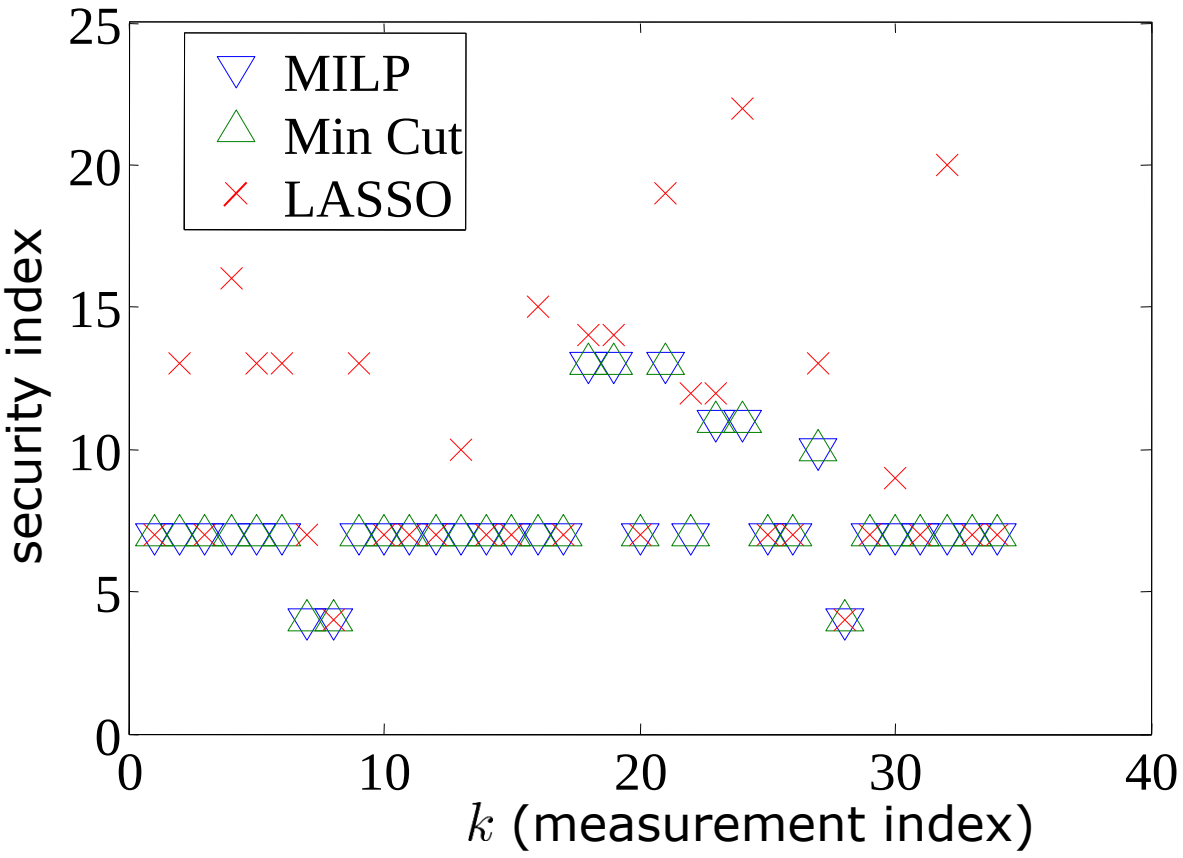
- Can we find solutions as **accurately** as MILP, and **faster** than LASSO?
  - Arbitrary *H*: **No**! (Problem NP-hard)
  - *H* with the special physical and measurement structure: **Yes**! (**Min Cut** polynomial time algorithm next)

- Can we find methods giving more **problem insight**, and ideas for **assigning protection**?
  - **Yes**, exploit graph interpretation of solution

IEEE 14 Bus Benchmark Test Result

Security indices for all measurements





Solve time: MILP 1.1s; LASSO 0.6s; Min Cut 0.02s

ROYAL INSTITUTE OF TECHNOLOGY

IEEE 118, 300, 2383 Bus Benchmarks

Min Cut solution is **exact**

Solve time comparison:

Method/Case	118 bus	300 bus	2383 bus
MILP	763 sec	6708 sec	About 5.7 days
Min Cut	0.3 sec	1 sec	31 sec

ROYAL INSTITUTE OF TECHNOLOGY

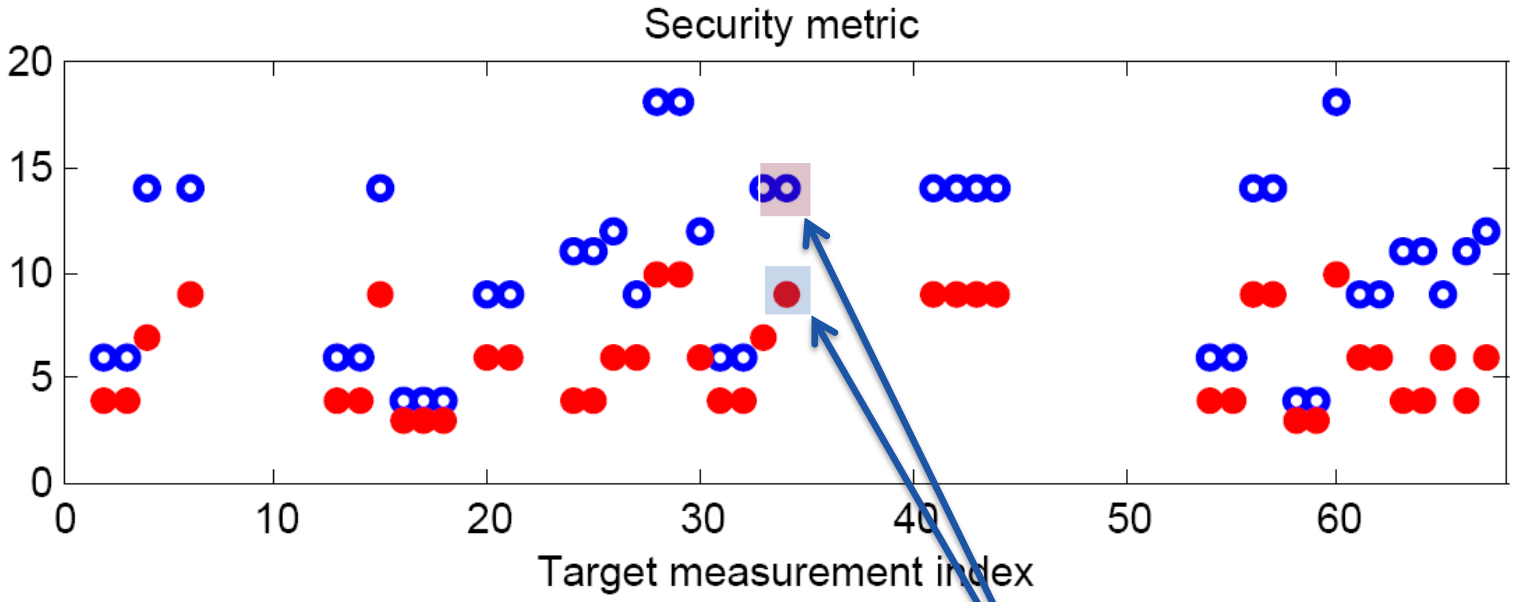
Wish List

- Can we find solutions as **accurately** as MILP, and **faster** than LASSO?
  - Arbitrary *H*: **No**! (Problem NP-hard)
  - *H* with the special physical and measurement structure: **Yes**! (Min cut polynomial time algorithm next.)

- Can we find methods giving more **problem insight**, and ideas for **assigning protection**?
  - **Yes**, exploit graph interpretation of solution
  - **Securing sensors that are frequently cut gives indirect protection to many sensors!**

    [Vukovic *et al.*, JSAC, 2012]

Assigning Protection Using Insight from Min Cut



- = Current measurement config.
- o = Upgraded measurement config.

With only a few sensors protected, 14 instead of 7 measurements has to be involved in a stealth attack!

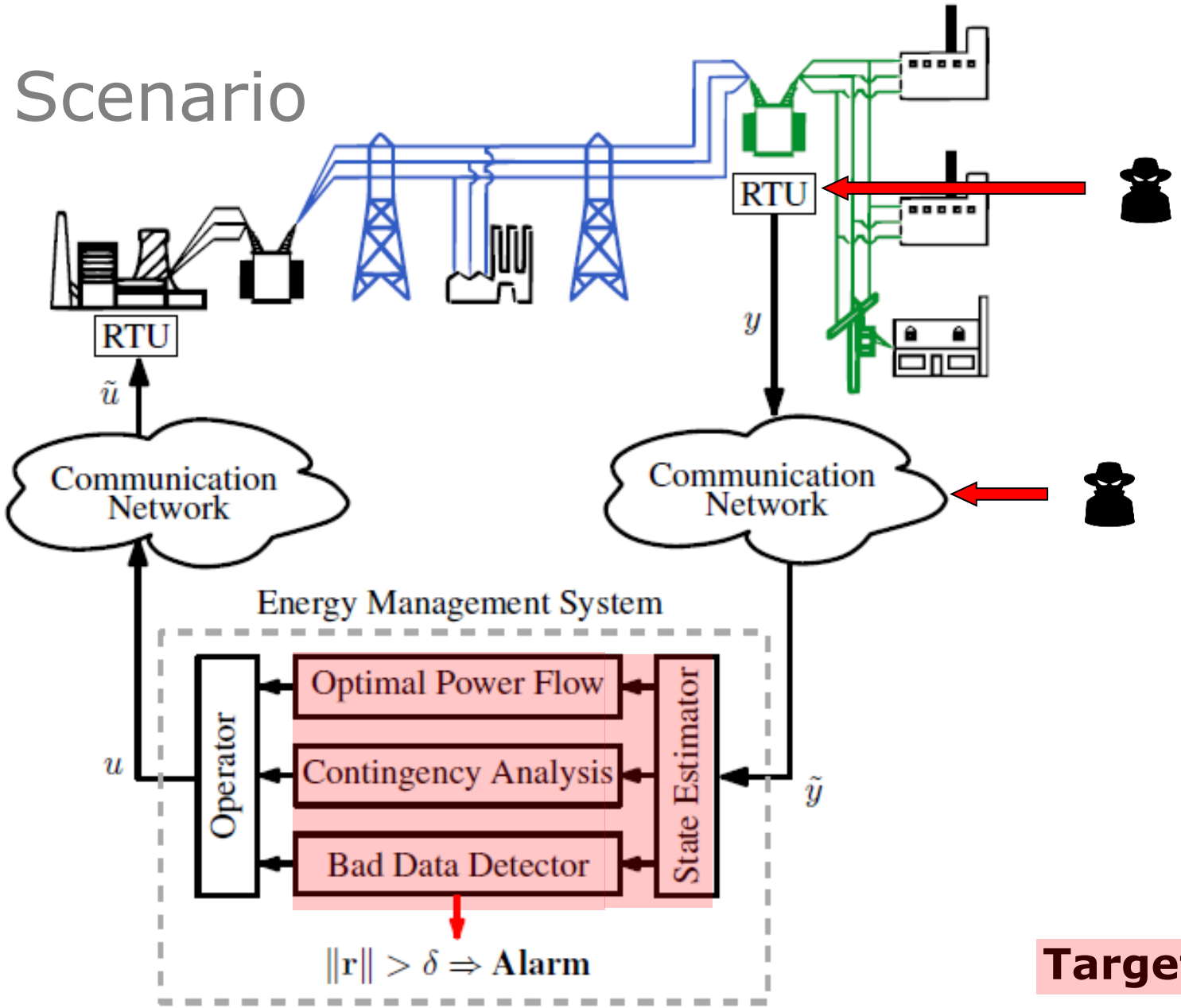
(Exercise 2)

The Scenario

Practically motivated problem…

How much security does the Bad Data Detector provide?

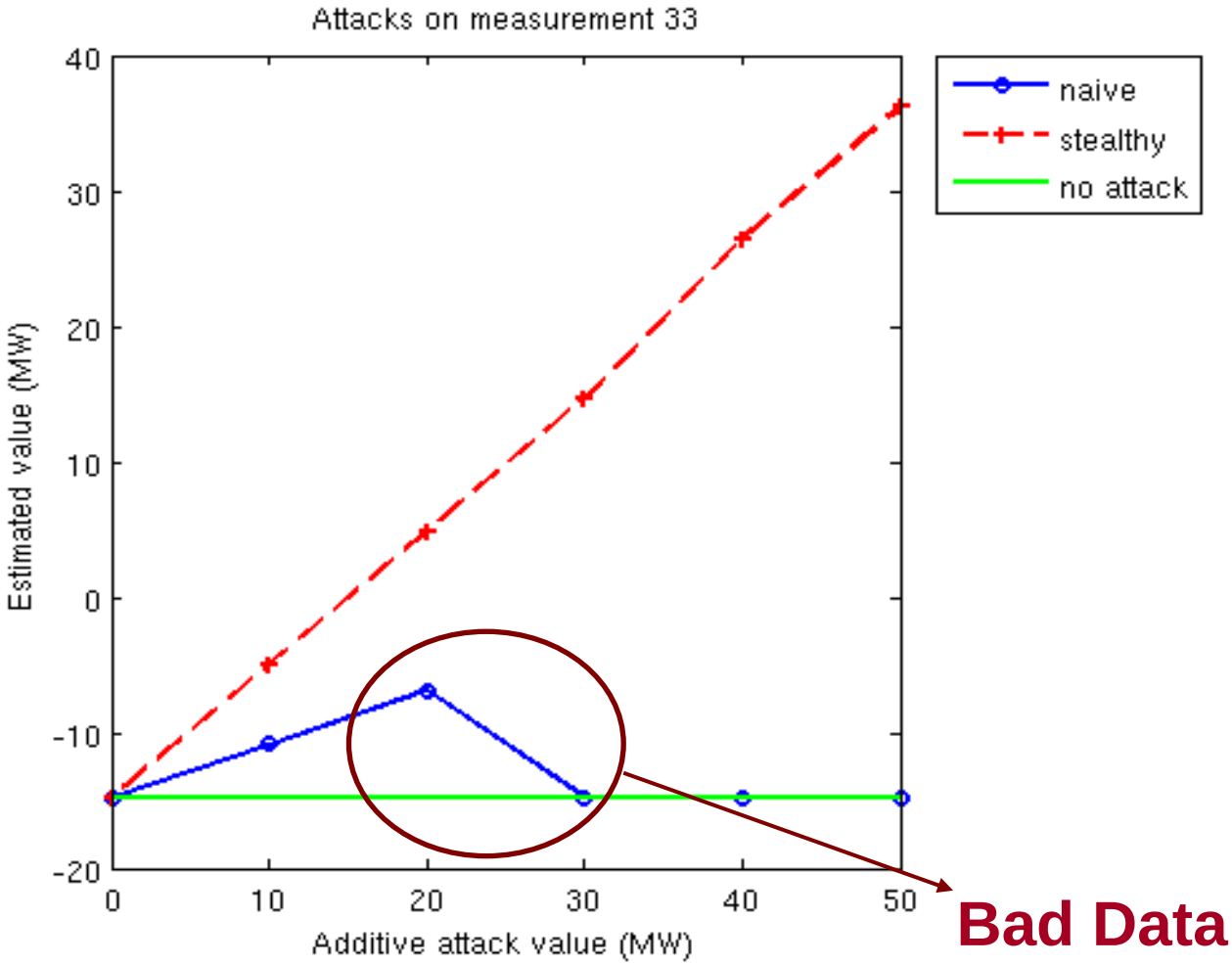
Some, when security index is high



Target

[Giani *et al.*, IEEE ISRCS, 2009]
[Mohajerin Esfahani *et al.*, CDC, 2010]

Verification in SCADA Testbed



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

Bad Data Detected & Removed

- Stealth attack of 150 MW (≈55% of nominal value) passed undetected in testbed!

[Teixeira *et al*., IFAC WC, 2011]

ROYAL INSTITUTE OF TECHNOLOGY

Lecture 1

- Background and motivation
- Adversaries in networked control systems
- Quantifying security: A case study in power system monitoring
- The security index, and its computation

Min Cut Algorithm on the Board

Compare:

"Hard":

$$\begin{aligned} \underset{\Delta\theta\in\mathbb{R}^{n+1}}{\text{minimize}} \quad & c^T g(DA^T\Delta\theta) + p^T g(ADA^T\Delta\theta) \\ \text{subject to} \quad & A(:,\bar{e})^T\Delta\theta \neq 0 \end{aligned} \tag{1}$$

"Easy":

$$\begin{aligned} \underset{\Delta\theta\in\{0,1\}^{n+1}}{\text{minimize}} \quad & c^T g(DA^T\Delta\theta) + p^T g(ADA^T\Delta\theta) \\ \text{subject to} \quad & A(:,\bar{e})^T\Delta\theta \neq 0. \end{aligned} \tag{2}$$

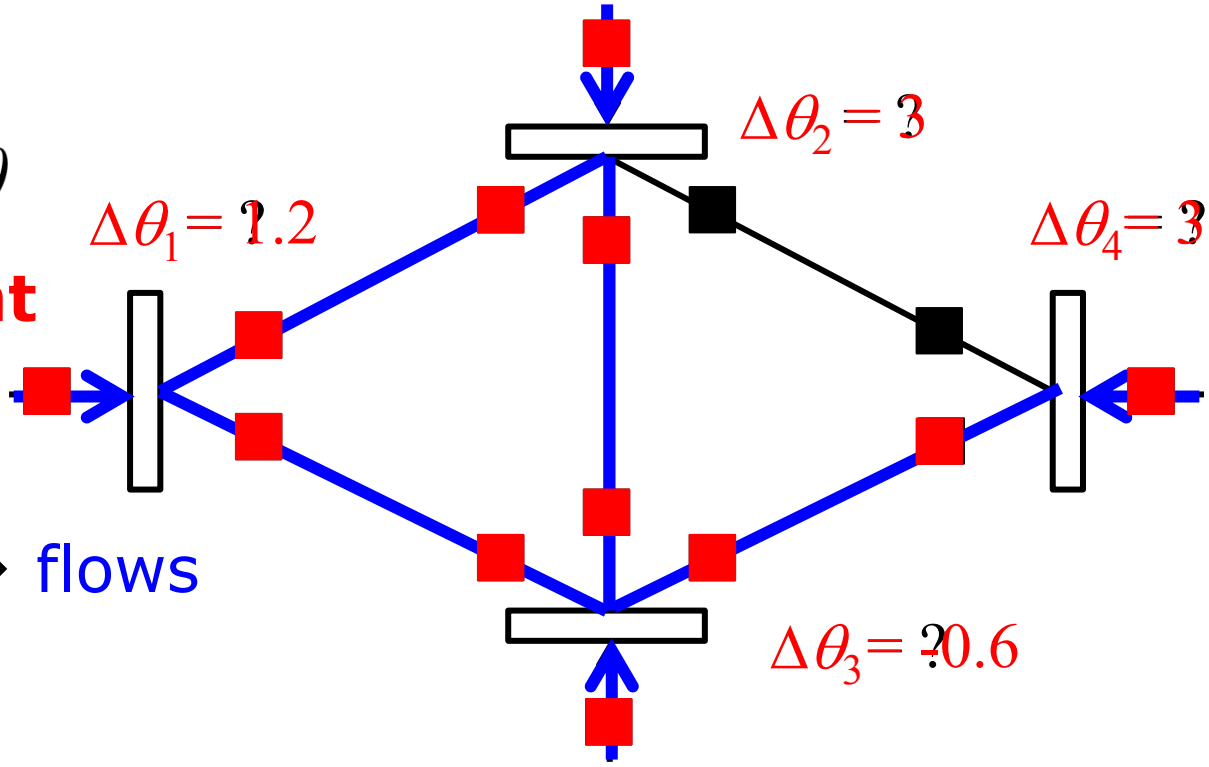
[Theorem 1, Hendrickx *et al*., IEEE TAC, 2014]

Graph Interpretation of Stealth Attack

Stealth attack $\Delta z = H\Delta\theta$

= **phase angle assignment**

Phase angle differences → flows



attack cost $\mathrm{card}(H\Delta\theta)$ = **# of meters with nonzero flows**

Binary Phase Assignment is Optimal

Security index problem

$$\min_{\Delta\theta} \operatorname{card}(H\Delta\theta)$$

$$\text{s.t.}$$

$$H(k,:)\Delta\theta = 1$$

[Sou *et al.*, CDC, 2011]

[Hendrickx *et al.*, IEEE TAC, 2014]

$$\min_{\Delta\theta} \operatorname{card}(H\Delta\theta)$$

$$\text{s.t.}$$

$$H(k,:)\Delta\theta = 1$$

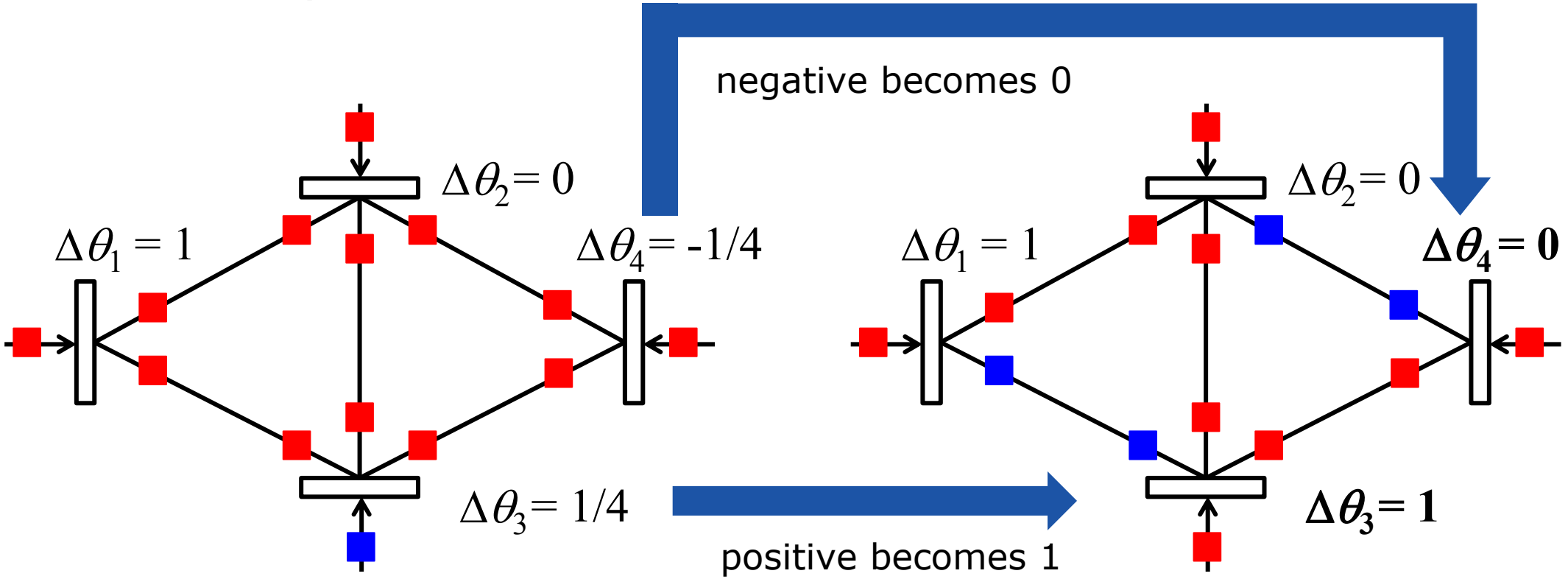
$$\Delta\theta_i \in \{0, 1\}$$

Theorem: Optimal $\Delta\theta_i$ can be restricted to 0 or 1, for all i

Proof: Restriction can never increase number of flows, given the structure of *H*

Binary Optimal Solution Justification

Can always find {0,1} feasible solution with no worst cost



Cost = **13** attacks

Cost = **10** attacks

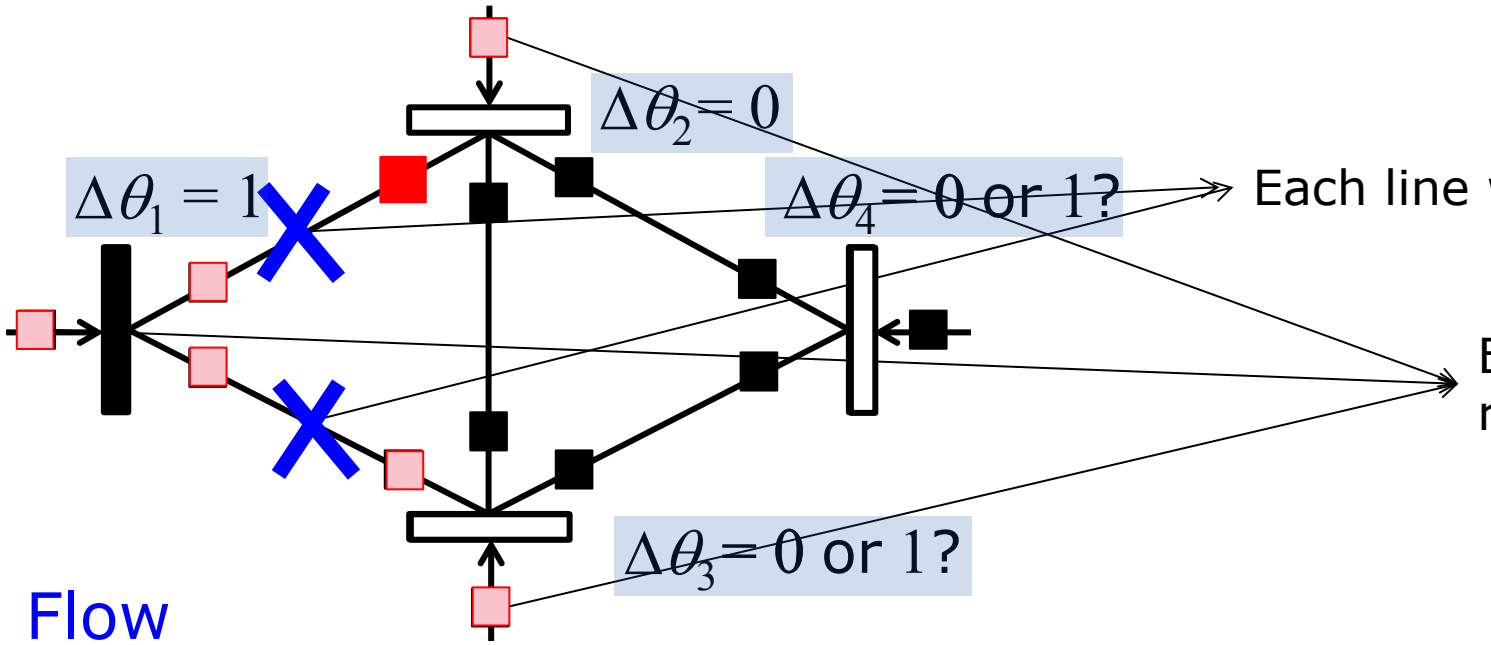
Reformulation as Node Partitioning

$$\min_{\Delta\theta} \operatorname{card}(H\Delta\theta)$$

$$\text{s.t.} \quad H(k,:)\Delta\theta = 1$$

$$\Delta\theta_i \in \{0,1\}$$

Security index problem:
Pick partition of minimum # of flows and incident nodes



Each line with flow requires 2 attacks

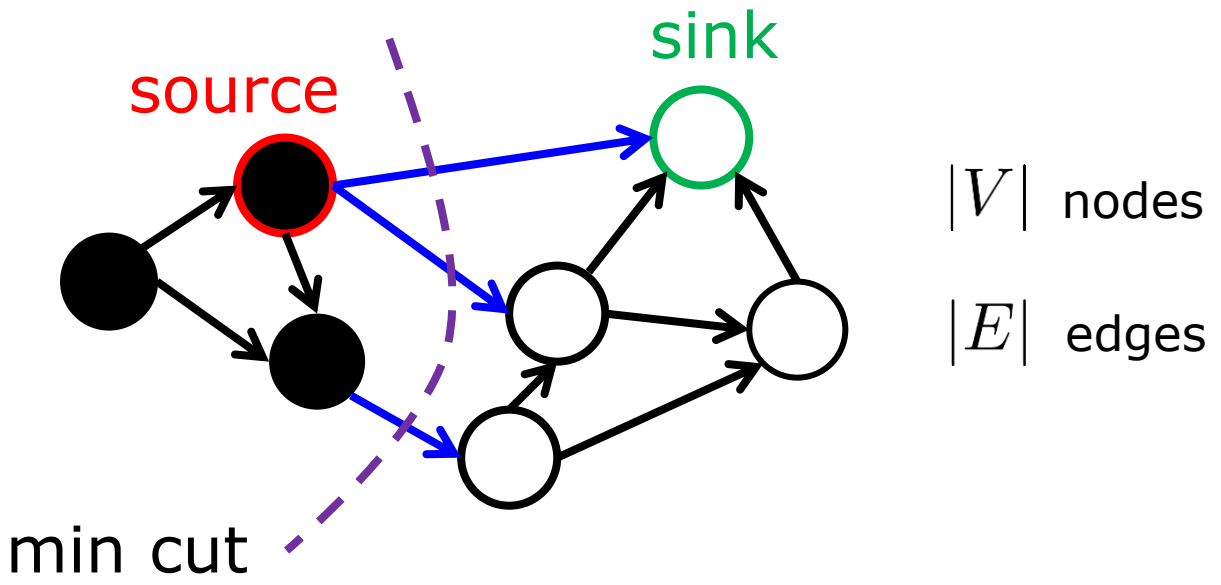
Each node incident requires 1 attack

Flow between partitions

target additional

ROYAL INSTITUTE OF TECHNOLOGY

Interlude: The Min Cut Problem

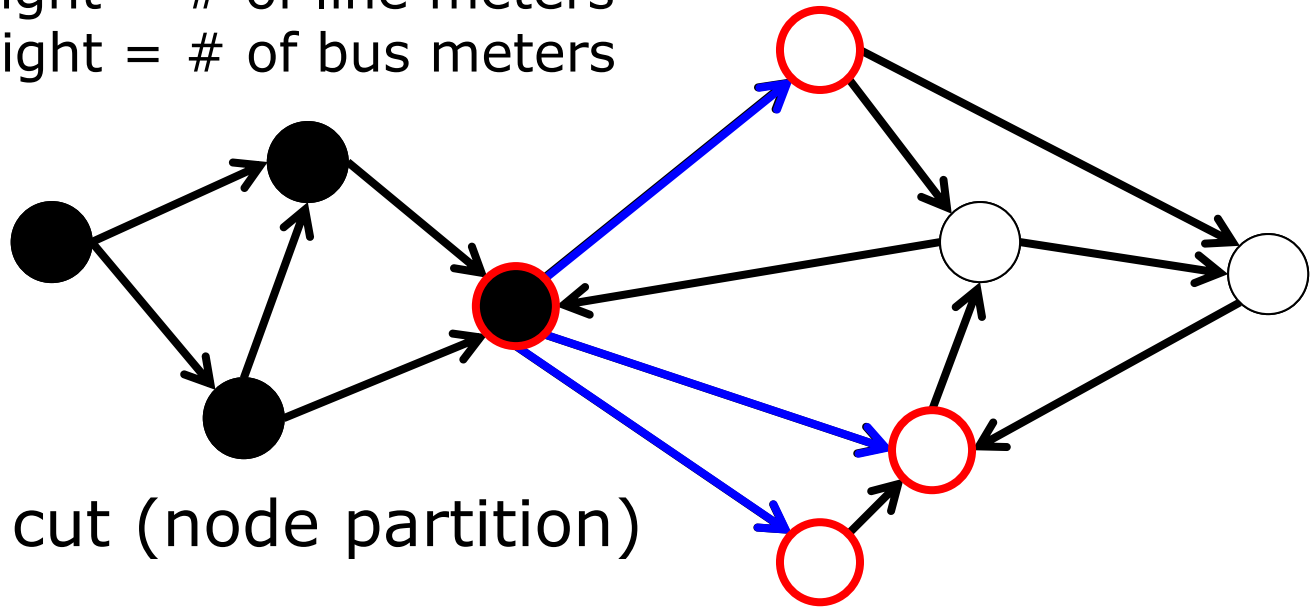


- Partition nodes into two sets (**black** and white) such that source is **black** and sink is white ("a cut")
- Find partitions with the smallest number of edges from source set to sink set ("a min cut")
- Problem solvable in $O(|V||E| + |V|^2 \log(|V|))$ operations

Generalized Min Cut with Costly Nodes

Focus on directed graph (undirected = bi-directed)

edge weight = # of line meters
node weight = # of bus meters



Find the cut (node partition)

to minimize weights of cut edge + incident node weights…

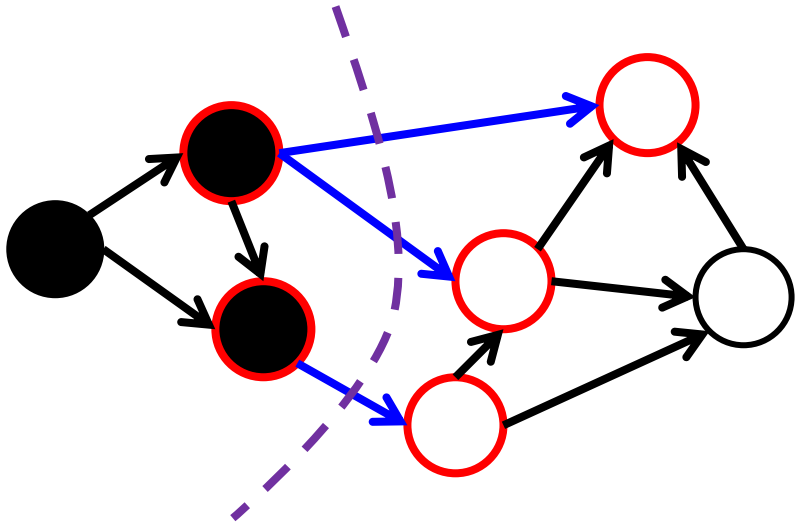
Generalization of standard Min Cut!

ROYAL INSTITUTE OF TECHNOLOGY

Security index problem

$$\min_{\Delta\theta \in \mathbb{R}^{n+1}} \ \mathrm{card}(H\Delta\theta)$$

$$\text{s.t. } \ H(k,:)\Delta\theta = 1$$



Generalized Min Cut problem

How to solve generalized Min Cut?

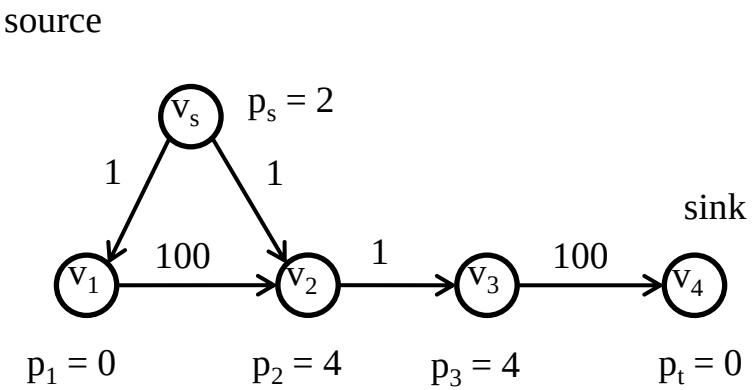
Standard Min Cut on Appended Graph

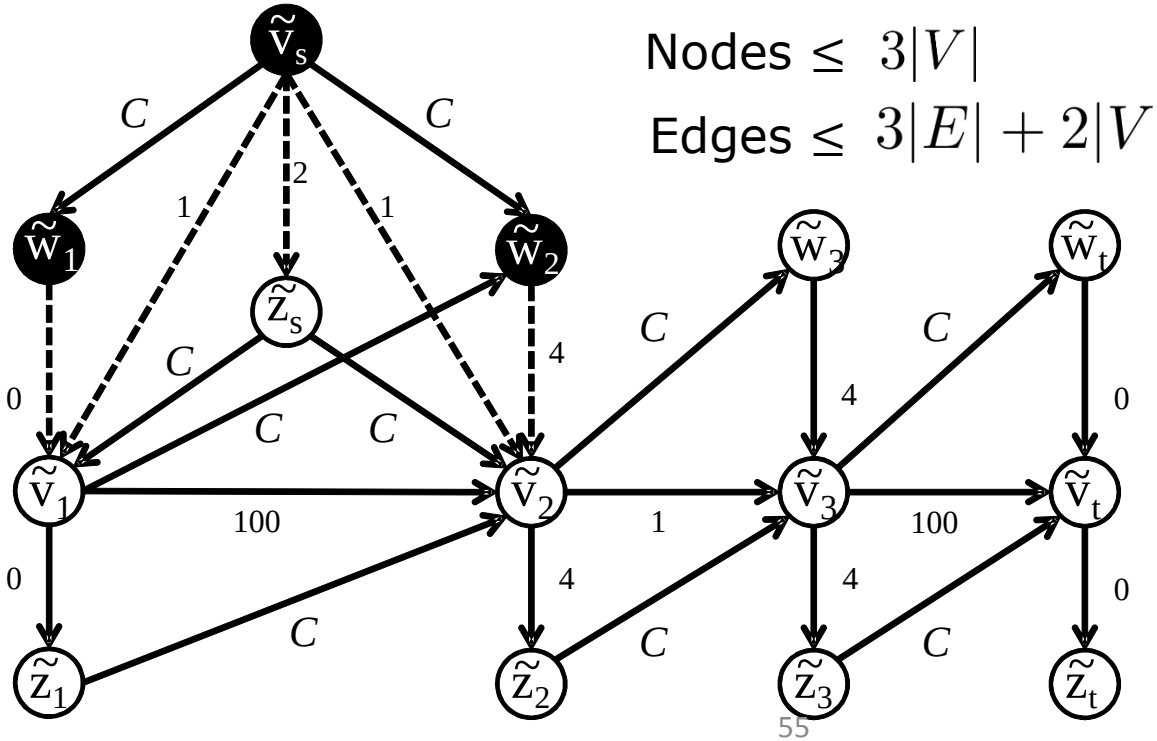
Generalized Min Cut = Standard Min Cut on **appended** graph

generalized min cut ⟺ standard min cut appended graph

$|V|$ nodes

$|E|$ edges





[Hendrickx *et al.*, IEEE TAC, 2014]

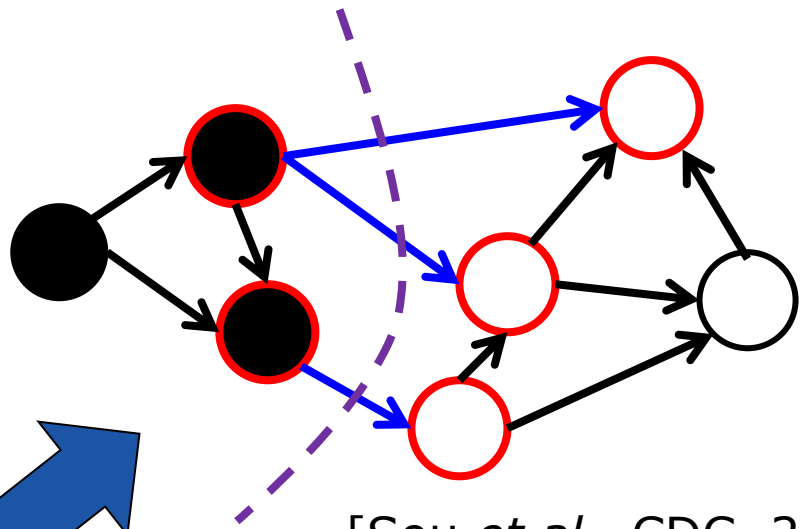
$$C > \max_i p_i$$

Security Index Problem – Summary

Security index problem

$$\min_{\Delta\theta} \operatorname{card}(H\Delta\theta)$$
$$\text{s.t. } H(k,:)\Delta\theta = 1$$

Generalized Min Cut problem



[Sou *et al*., CDC, 2011]

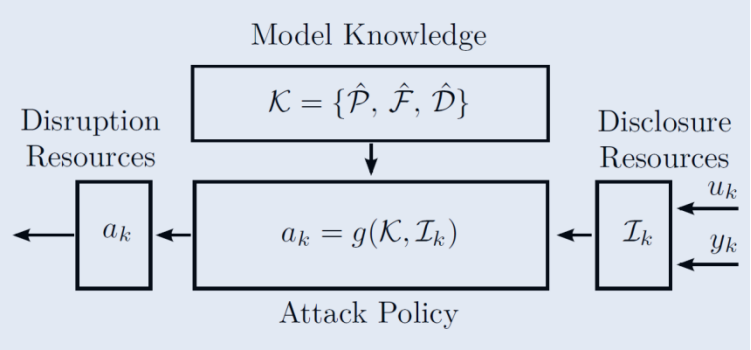
[Hendrickx *et al*., TAC, accepted 2014]

Standard Min Cut problem on an **appended** graph

```
>> [maxflow,mincut] = max_flow(A,source,sink);
```

Summary

- Cyber threats in control systems
- Adversary model in control systems



- A security metric for state estimators and its computation using Min Cut

$$\min_{\Delta\theta} \operatorname{card}(H\Delta\theta)$$
$$\text{s.t. } H(k,:)\Delta\theta = 1$$

- Exercises: Compute and understand the security metric
- Tomorrow: Attack space for cyber-physical systems