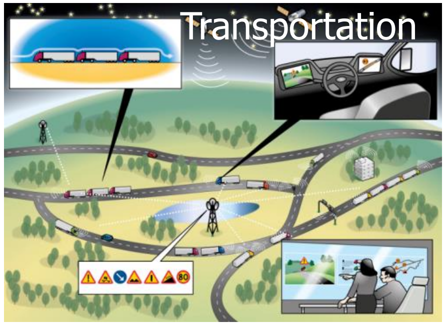
KTH ROYAL INSTITUTE
OF TECHNOLOGY

From Control System Security Indices to Attack Identifiability

Henrik Sandberg
KTH Automatic Control
hsan@kth.se

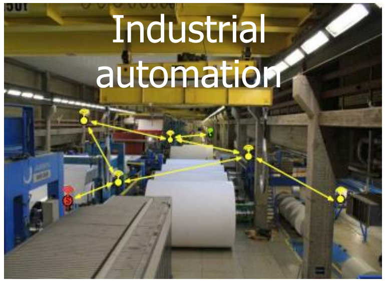
André Teixeira
Delft University of Technology
andre.teixeira@tudelft.nl

Motivation



Complex control systems with numerous attack scenarios

Examples: Critical infrastructures (power, transport, water, gas, oil) often with weak security guarantees

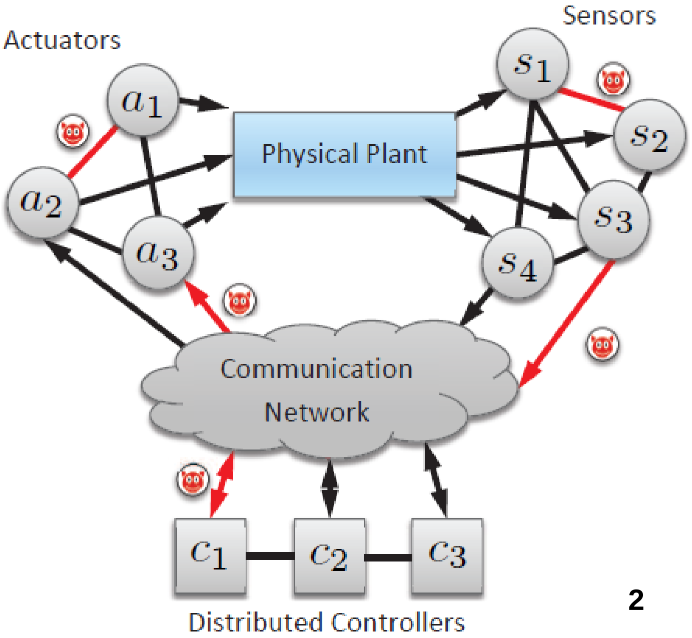


Too costly to secure the entire system against all attack scenarios

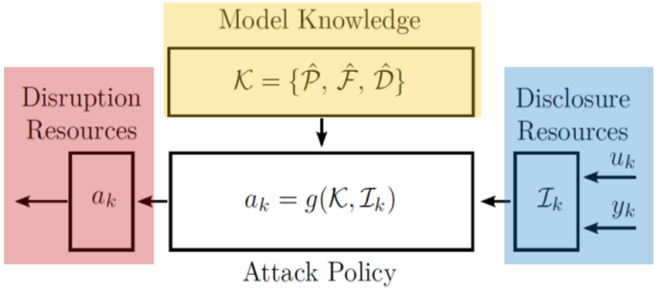
What scenarios to prioritize?

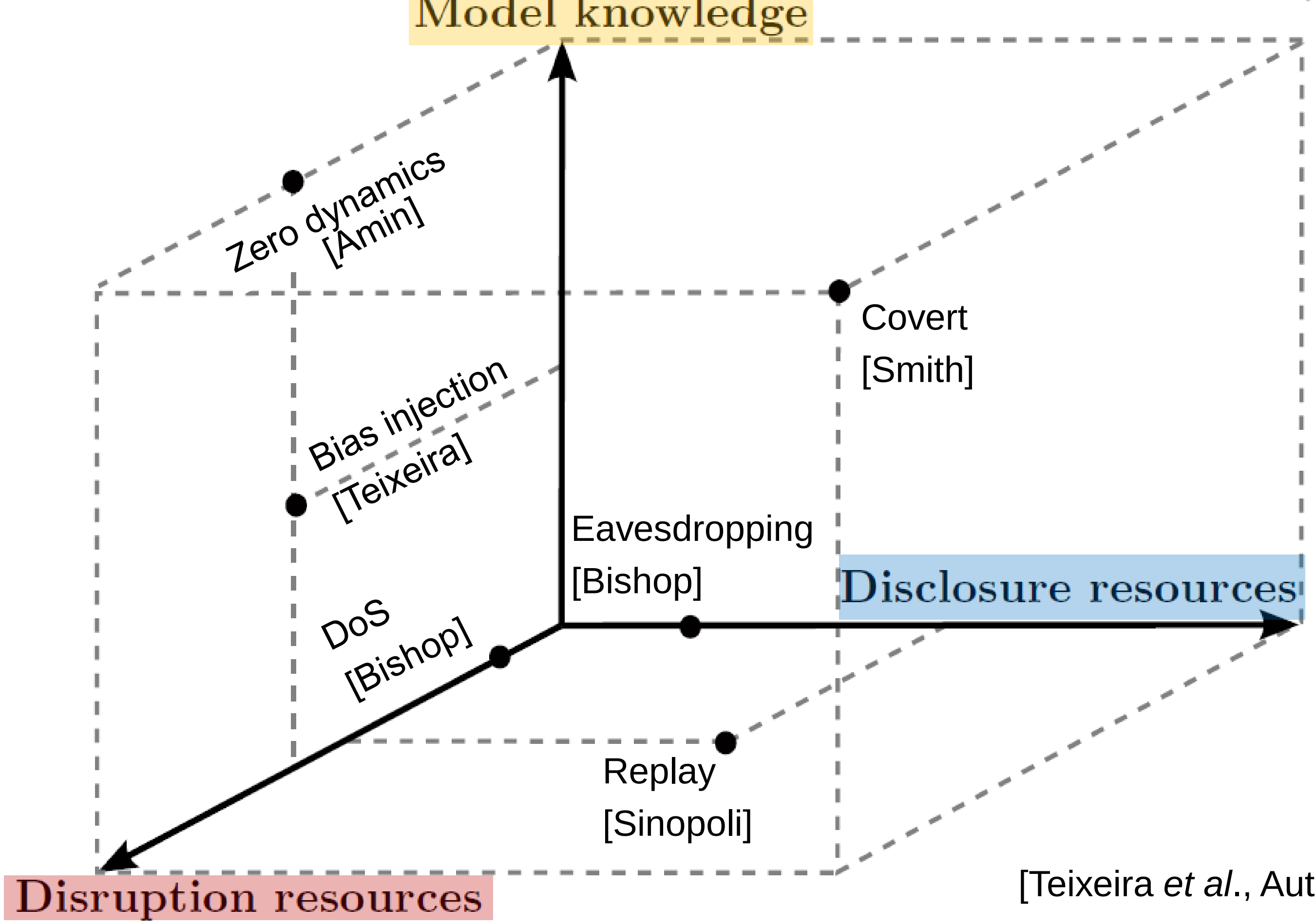
What components to protect?

When possible to identify attacks?



Control Systems Attack Space


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



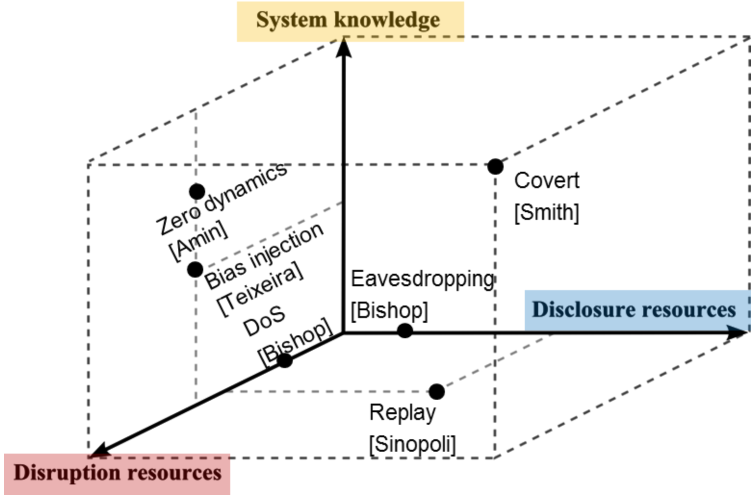
[Teixeira *et al*., Automatica, 2015]

Outline

- **Risk management**
- Dynamical security index
- Special cases and computational issues
  - Critical signals
  - Transmission zeros
  - Sensor attacks
  - Static systems
- Attack identification and secure state estimation

Defining Risk

Risk = (Scenario, Likelihood, Impact)

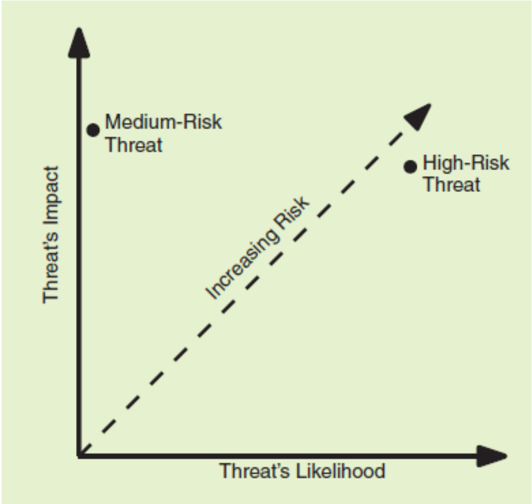


Scenario

- How to describe the system under attack?

Likelihood

- How much effort does a given attack require?



Impact

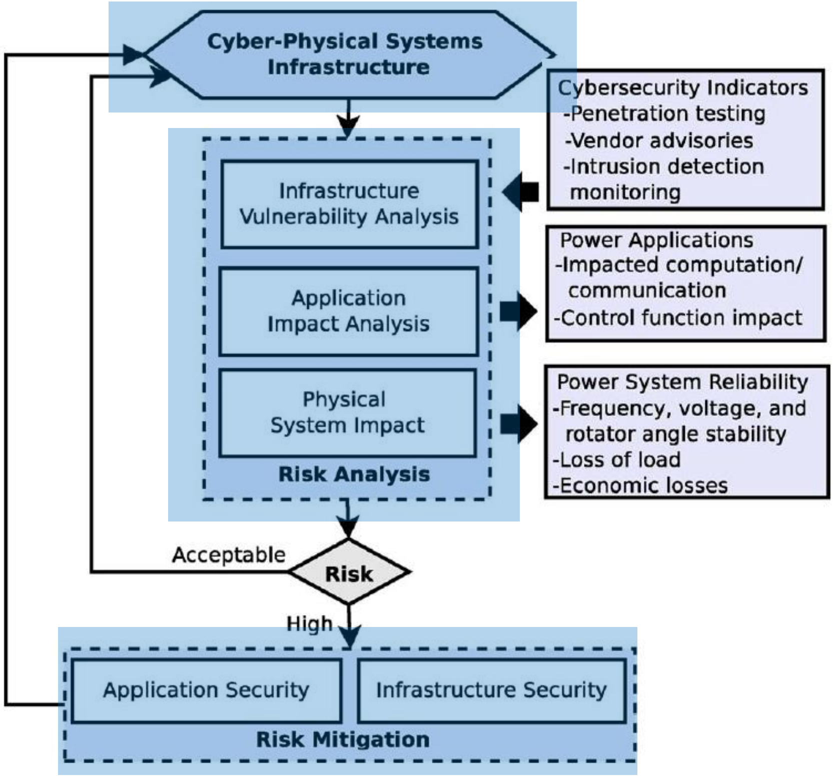
- What are the consequences of an attack?

[Kaplan & Garrick, 1981], [Bishop, 2002]
([Teixeira *et al*., IEEE CSM, 2015])

Risk Management Cycle

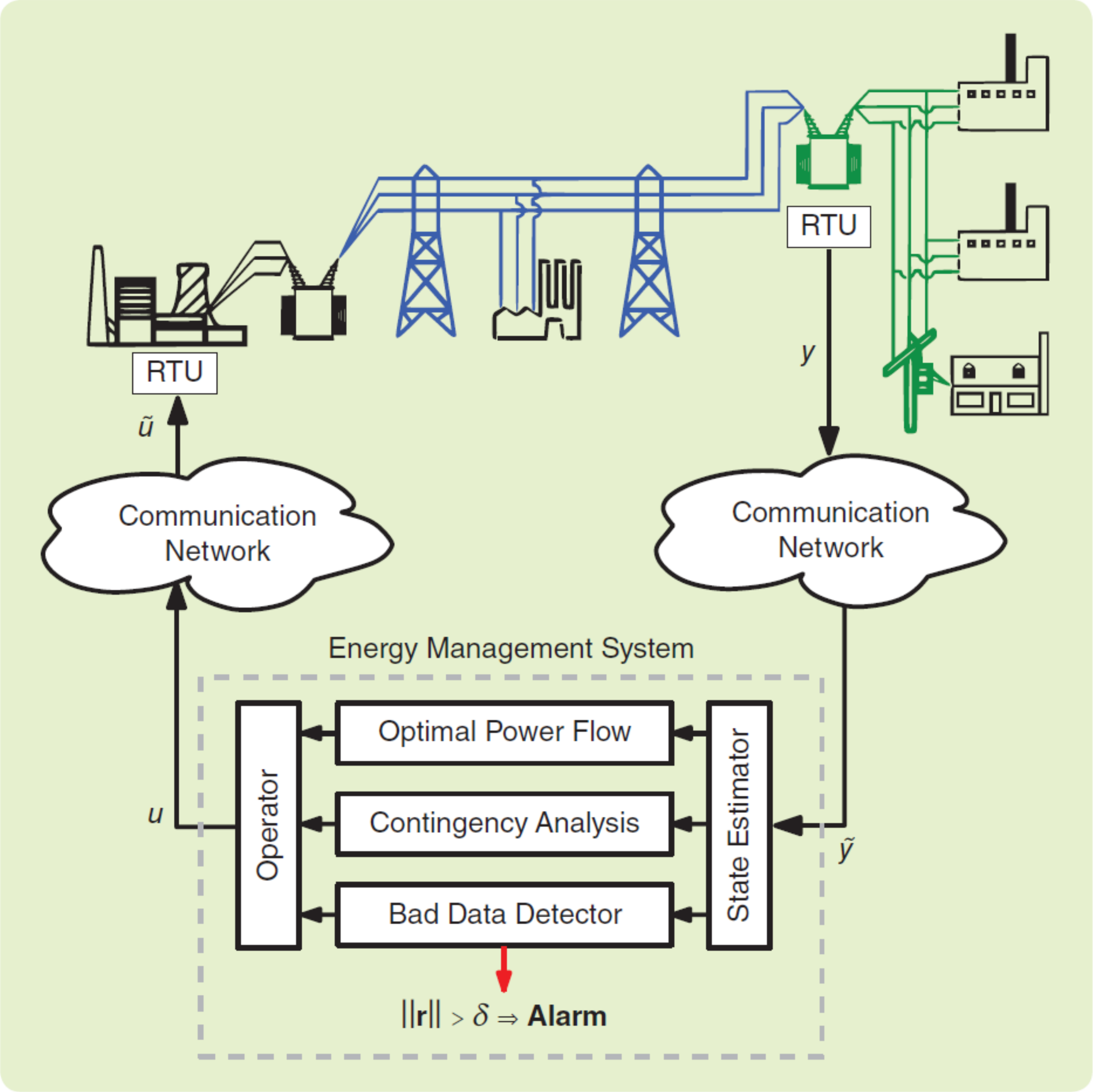
Main steps in risk management

- Scope definition
  - Models, Scenarios, Objectives
- Risk Analysis
  - Threat Identification
  - **Likelihood Assessment**
  - Impact Assessment
- Risk Treatment
  - Prevention, **Detection**, Mitigation

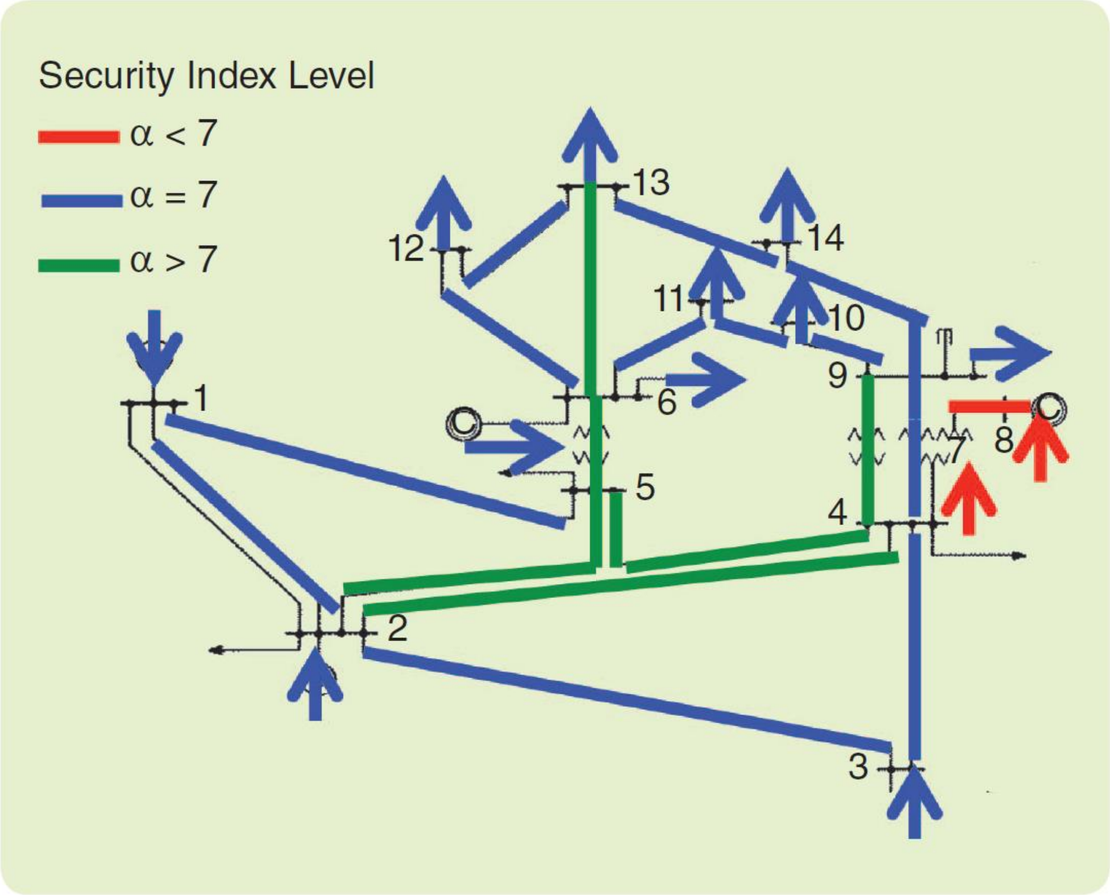


[Sridhar *et al*., Proc. IEEE, 2012]

Example: Power System State Estimator



Example: Power System State Estimator



Security index α (to be defined) indicates sensors with inherent weak redundancy (~security). These should be secured first!

[Teixeira *et al*., IEEE CSM, 2015], [Vukovic *et al.*, IEEE JSAC, 2012]

Outline

- Risk management
- **Dynamical security index**
- Special cases and computational issues
    - Critical signals
    - Transmission zeros
    - Sensor attacks
    - Static systems
- Attack identification and secure state estimation

Model and Definitions

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

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

- Unknown state $x(k) \in \mathbb{R}^n$
- Unknown (natural) disturbance $d(k) \in \mathbb{R}^o$
- Unknown (malicious) attack $a(k) \in \mathbb{R}^m$
- Known measurement $y(k) \in \mathbb{R}^p$
- Known model A, B_d, B_a, C, D_d, D_a

- **Definition:** Attack signal a is *persistent* if $a(k) \nrightarrow 0$ as $k \to \infty$

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

Undetectable Attacks and Masking

The Rosenbrock system matrix:

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

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

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

- Attack signal is undetectable if indistinguishable from measurable (y) effects of natural noise (d_0) or uncertain initial states (x_0) [**masking**]

- Compare with *fault detection* set-up with (*non-malicious*) faults and natural disturbances

The Security Index α_i

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

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

Interpretation:

- Attacker persistently targets element a_i (condition $|z_0| \geq 1$)
- α_i is smallest number of attack signals that need to be simultaneously accessed for undetectability

Argument: Large $\alpha_i \Rightarrow$ malicious cyber attacks targeting a_i less likely

Problem NP-hard in general (combinatorial optimization, cf. matrix *spark*). Generalization of static index in [Sandberg *et al.*, SCS, 2010]

Special Case 1: Critical Attack Signals

Signal with $\alpha_i = 1$ can be undetectably attacked without access to other elements ⇒ **Critical Attack Signal**

$$P_i(z) = \begin{bmatrix} A - zI & B_d & B_{a,i} \\ C & D_d & D_{a,i} \end{bmatrix} \in \mathbb{C}^{(n+p)\times(n+o+1)}, \; P_d(z) = \begin{bmatrix} A - zI & B_d \\ C & D_d \end{bmatrix} \in \mathbb{C}^{(n+p)\times(n+o)}$$

Simple test, $\forall \boldsymbol{i}$: If there is $z_0 \in \mathbb{C}$, $|z_0| \geq 1$, such that $\text{rank}\,[P_d(z_0)] = \text{rank}\,[P_i(z_0)]$, then $\alpha_i = 1$

Even more critical case: If $\text{normalrank}\,[P_d(z_0)] = \text{normalrank}\,[P_i(z_0)]$ then there is undetectable critical attack for all frequencies z_o

Holds generically when more disturbances than measurements ($o \geq p$)!

Secure against these attack signals first in risk management!

Special Case 2: Transmission Zeros

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

[Amin *et al.*, ACM HSCC, 2010]
[Pasqualetti *et al.*, IEEE TAC, 2013]

Suppose $P(z)$ has full column normal rank. Then undetected attacks only at finite set of transmission zeros $\{z_0\}$

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

by inspection of corresponding zero directions ⇒ **Easy in typical case of 1-dimensional zero directions**

Special Case 3: Sensor Attacks

$$P(z) = \begin{bmatrix} A - zI & 0 & 0 \\ C & D_d & D_a \end{bmatrix}$$

[Fawzi *et al*., IEEE TAC, 2014]
[Chen *et al*., IEEE ICASSP, 2015]
[Lee *et al*., ECC, 2015]

$P(z)$ only loses rank in eigenvalues $z_0 \in \{\lambda_1(A), \ldots, \lambda_n(A)\}$

Simple eigenvalues give one-dimensional spaces of eigenvectors x_0 ⇒ **Simplifies computation of $\boldsymbol{\alpha_i}$**

Example: Suppose $D_a = I_p$ (sensor attacks), $D_d = 0$, and system observable from each y_i, $i = 1, \ldots, p$:

- By the PBH-test: $\alpha_i = p$ or $\alpha_i = +\infty$ (if all stable eigenvalues, no persistent undetectable sensor attack exists)
- Redundant measurements increase α_i!

Special Case 4: Sensor Attacks for Static Systems

$$P(z) = \begin{bmatrix} I - zI & 0 & 0 \\ C & 0 & D_a \end{bmatrix}$$

[Liu *et al.*, ACM CCS, 2009]
[Sandberg *et al.*, SCS, 2010]

Since $A = I_n$ and $B_d = B_a = 0$, this is the steady-state case

Space of eigenvectors x_0 is n-dimensional ⇒ **Typically makes computation of α_i harder than in the dynamical case!**

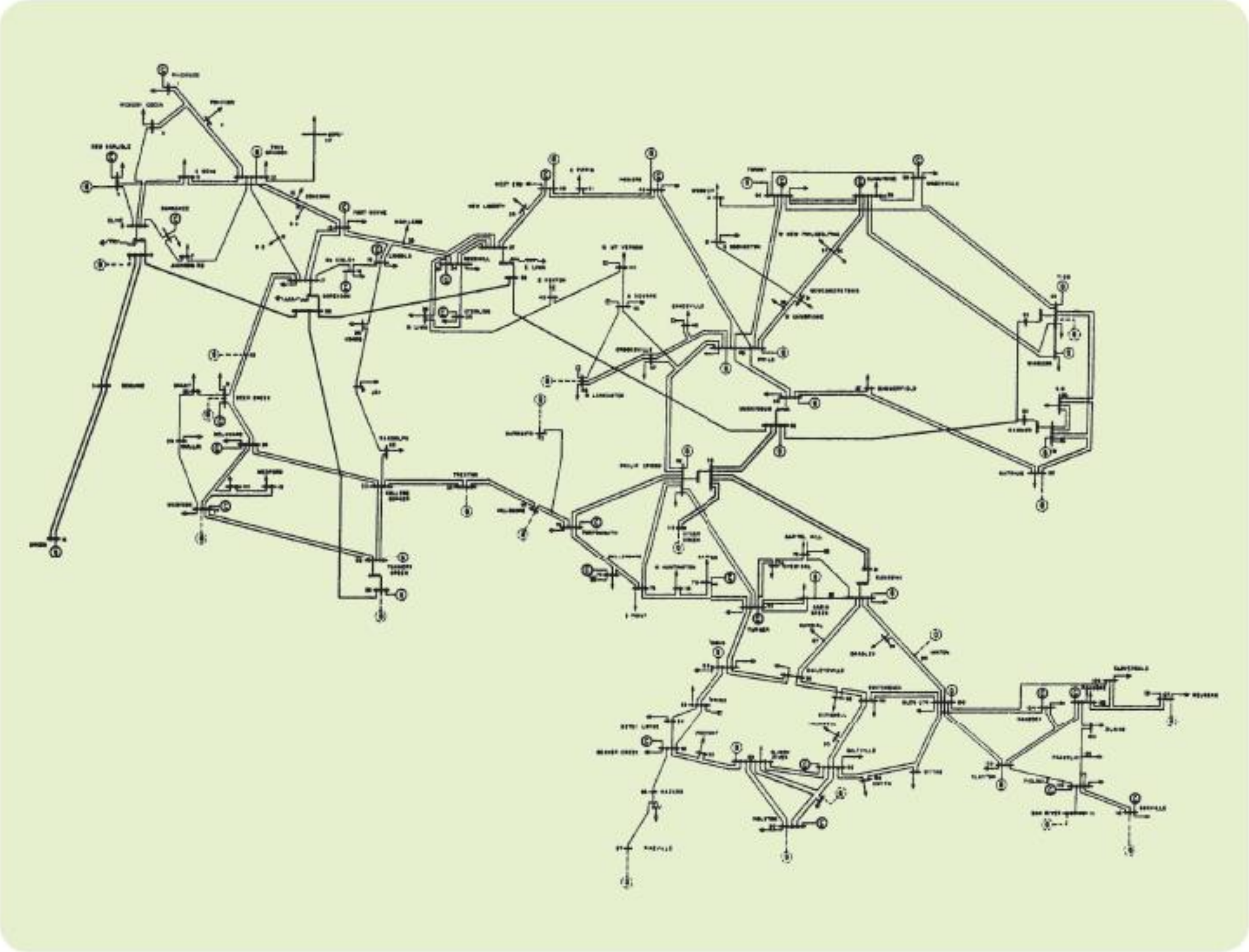
Practically relevant case in power systems where $p > n \gg 0$

- Problem NP-hard, but power system imposes special structures in C (unimodularity etc.)
- Several works on efficient and exact computation of α_i using min-cut/max-flow and ℓ_1-relaxation ([Hendrickx *et al*., 2014], [Kosut, 2014], [Yamaguchi *et al*., 2015])

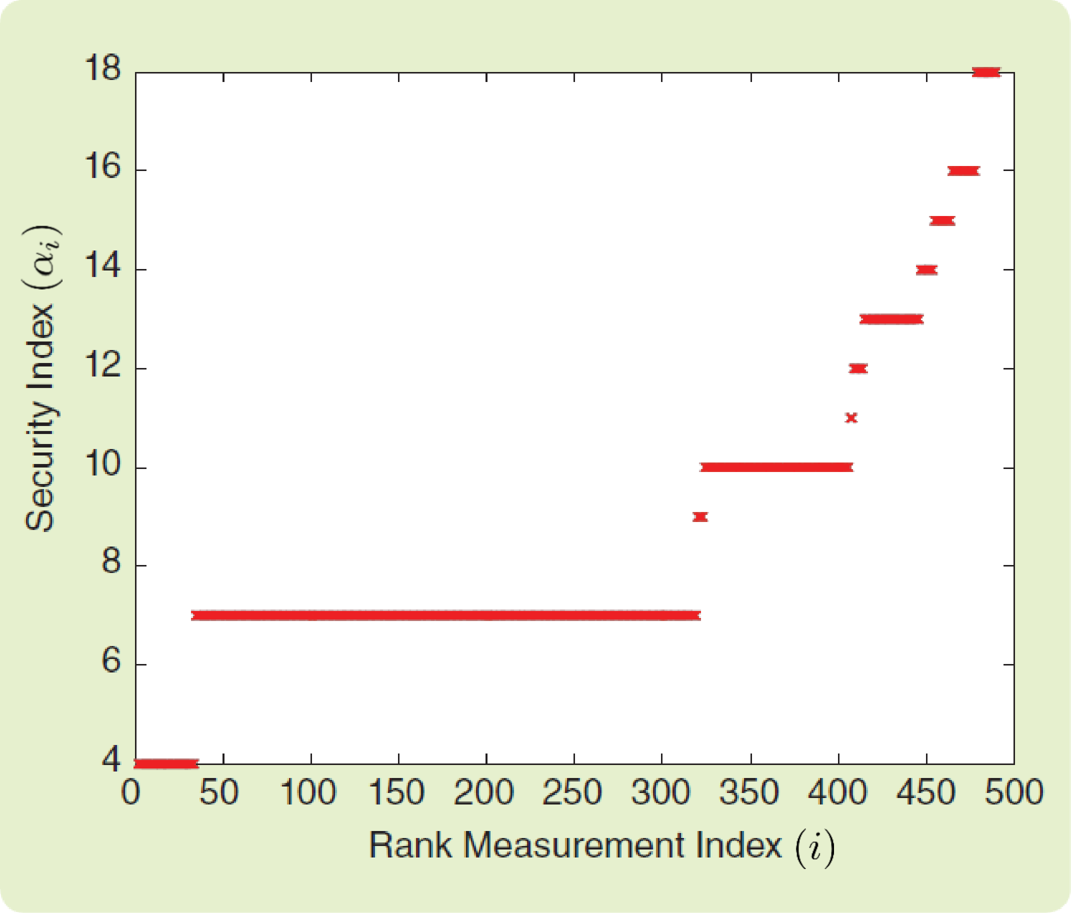
Example: Power System State Estimator for IEEE 118-bus System

- State dimension $n = 118$

- Number sensors $p \approx 490$



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



- Computation time on laptop using min-cut method [Hendrickx *et al.*, IEEE TAC, 2014]: 0.17 sec
- Note the wide spread of indices. Greedy method for security allocation used in [Vukovic *et al.*, IEEE JSAC, 2012]

Summary So Far

- Dynamical security index α_i defined

- Argued α_i useful in risk management for assessing likelihood of malicious attack against element a_i

- Computation is NP-hard in general, but often “simple” in special cases:
  - One-dimensional zero-dynamics
  - Static systems with special matrix structures (derived from potential flow problems)
  - Dynamics generally simplifies computation and redundant sensors increase α_i

- Fast computation enables greedy security allocation

Outline

- Risk management
- Dynamical security index
- Special cases and computational issues
  - Critical signals
  - Transmission zeros
  - Sensor attacks
  - Static systems
- **Attack identification and secure state estimation**

Attack Identification

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

- Unknown state $x(k) \in \mathbb{R}^n$
- Unknown (natural) disturbance $d(k) \in \mathbb{R}^o$
- Unknown (malicious) attack $a(k) \in \mathbb{R}^m$
- Known measurement $y(k) \in \mathbb{R}^p$
- Known model A, B_d, B_a, C, D_d, D_a

- When can we decide there is an attack signal $a_i \neq 0$?
- Which elements a_i can we track (“identify”)?

- Not equivalent to designing an unknown input observer/secure state estimator (state not requested here). See end of presentation

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 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
- 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, $j \neq i$

Theorem

Suppose that the attacker can manipulate at most q attack elements simultaneously ($\|a\|_0 \leq q$).

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

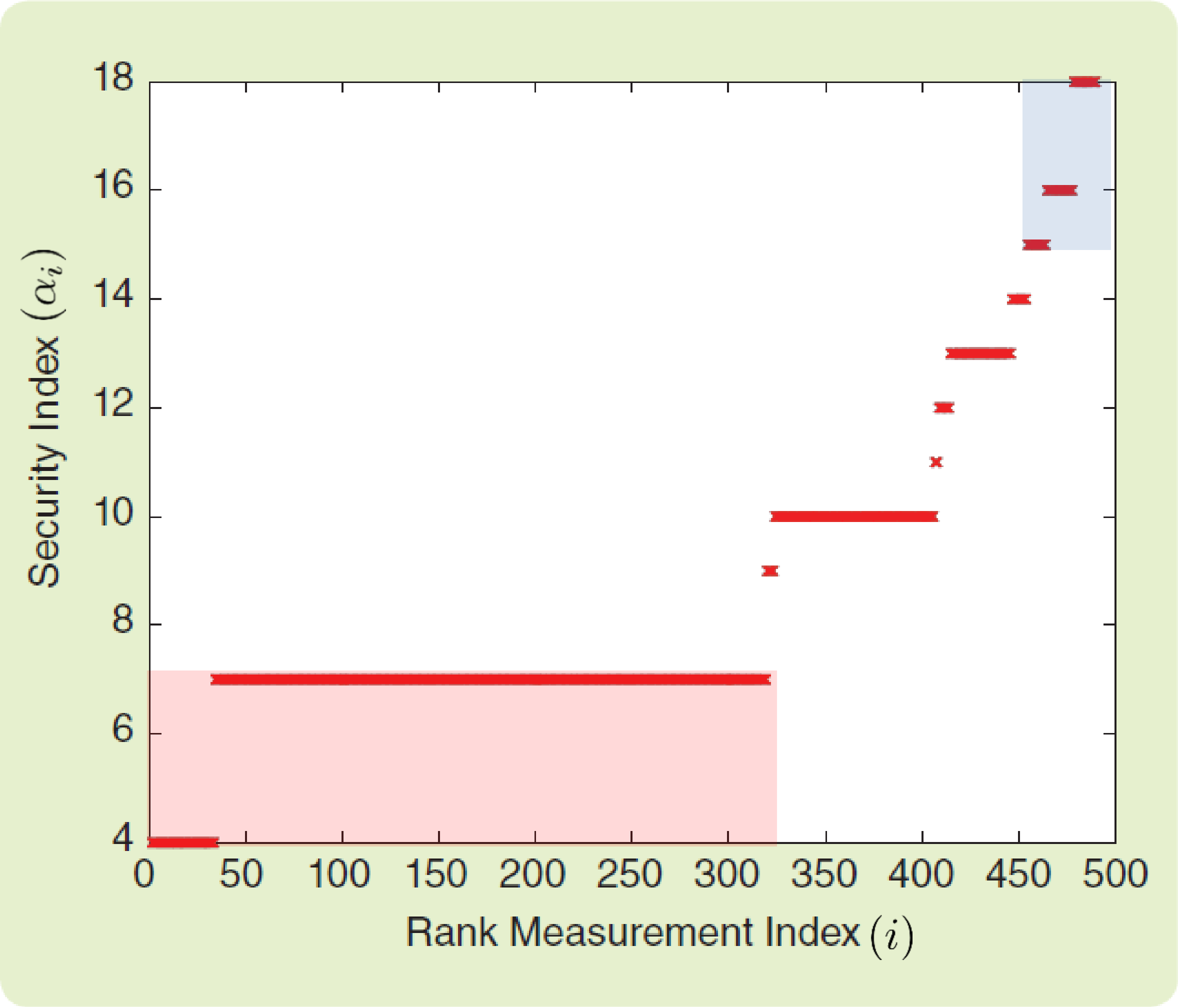
ii. All persistent attacks are i-identifiable iff $q < \alpha_i/2$;

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

Proof. Compressed sensing type argument. See paper for details

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

- Suppose number of attacked elements is $q \leq 7$



- Signals susceptible to undetectable attacks

- Signals were all attacks are identifiable

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

Secure State Estimation/Unknown Input Observer (UIO)

Secure state estimate $\hat{x}$: Regardless of disturbance d and attack a, the estimate satisfies $\hat{x} \to x$ as $k \to \infty$

1. Rename and transform attacks and disturbances:

$$\begin{bmatrix} B_d \\ D_d \end{bmatrix} d + \begin{bmatrix} B_a \\ D_a \end{bmatrix} a = \begin{bmatrix} B_f \\ D_f \end{bmatrix} f, \quad \text{such that} \begin{bmatrix} B_f \\ D_f \end{bmatrix} \text{full column rank}$$

2. Compute security indices α_i with respect to f

Theorem: A secure state estimator exists iff

1. (C, A) is detectable; and

2. $q < \min_i \frac{\alpha_i}{2}$, where q is max number of non-zero elements in f.

Proof. Existence of UIO by [Sundaram *et al.*, 2007] plus previous theorem

How to Identify an Attack Signal?

Use decoupling theory from fault diagnosis literature [Ding, 2008]

Suppose that $y = G_d d + G_a a$ and

$$\begin{aligned}\text{normalrank}\,[G_d(z)] &= m',\\ \text{normalrank}\,[G_d(z)\ G_a(z)] &= m' + m''\end{aligned}$$

Then there exists linear decoupling filter R such that

$$\begin{bmatrix} r \\ y' \end{bmatrix} = R(G_d d + G_a a) = \begin{bmatrix} 0 & \Delta \\ G'_d & G'_a \end{bmatrix} \begin{bmatrix} d \\ a \end{bmatrix},$$

$$\text{normalrank}\,[G'_d(z)] = \text{normalrank}\,[G'_d(z)\,G'_a(z)] = m'$$

$$\text{normalrank}\,[\Delta(z)] = m''$$

How to Identify an Attack Signal?

Suppose a is identifiable ($q < \min_i \alpha_i/2$)

1. Decouple the disturbances to obtain system $r = \Delta a$

2. Filter out uncertain initial state component in r to obtain $r' = \Delta a$

3. Compute left inverses of $\Delta_I := [\Delta_i]_{i\in I}$ formed out of the columns Δ_i of Δ, for all subsets $|I| = \ q, I \subseteq \ \{1, \dots, m\}$ (**Bottleneck! Compare with compressed sensing**)

4. By identifiability, if estimate $\hat{a}_I$ satisfies $r' = \Delta\hat{a}_I$, then $\hat{a}_I \equiv a$

(Similar scheme applies if a is only i-identifiable)

Summary

- Dynamical security index α_i was defined and computational issues were raised
- Suppose attacker has access to q elements:
  - Undetectable attacks against a_i iff $q \geq \alpha_i$
  - Attack against a_i identifiable iff $q < \alpha_i/2$
- Argued α_i is useful in risk management for assessing likelihood of malicious attack against element a_i
- Many useful results in the fault diagnosis literature, especially for detectable attacks
- Research direction: More accurate attacker models, inspired by systems security