Control Systems Security Metrics and Risk Management

Henrik Sandberg

hsan@kth.se

Department of Automatic Control, KTH, Stockholm, Sweden

DISC Summer School, The Hague, The Netherlands, July 5
Example of Classic Cyber Security: The Byzantine Generals Problem

- Consider $n$ generals and $q$ unknown traitors among them. Can the $n - q$ loyal generals always reach an agreement?
- Traitors (“Byzantine faults”) can do anything: different message to different generals, send no message, change forwarded message,…
- Agreement protocol exists iff $n \geq 3q + 1$
- If loyal generals use unforgeable signed messages (“authentication”) then agreement protocol exists for any $q$! [Lamport et al., ACM TOPLAS, 1982]
- Application to linear consensus computations: See [Pasqualetti et al., CDC, 2007], [Sundaram and Hadjicostis, ACC, 2008]
Observations and Goals of Lecture

(q used as general proxy for overall attacker strength in the following)

- Resourceful attacker (large q) is hard/impossible to stop
- Actual q probably not known – Use varying q as input to risk study
- Large-scale industrial control systems are relatively unprotected today - Even small q may lead to substantial damage
- Smart defense can (significantly) increase the attacker’s required q

Goals of lecture

- Introduction to risk management and attack space
- Find signals susceptible to undetectable/unidentifiable attacks as fcn of q
- Introduce security metric (index) $\alpha$ and its computation
- Method to allocate defense to increase attacker’s required q
Outline

• Risk management

• Attack detectability and security metric

• Attack identification and secure state estimation

• Security metric 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
  \[
  \tilde{u}_k = u_k + \Gamma^u b^u_k \quad \tilde{y}_k = y_k + \Gamma^y b^y_k
  \]
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

\[ a_k = g(\mathcal{K}, \mathcal{I}_k) \]

Model Knowledge

\[ \mathcal{K} = \{ \hat{\mathcal{P}}, \hat{\mathcal{F}}, \hat{\mathcal{D}} \} \]

Disruption Resources

\[ a_k \]

Disclosure Resources

\[ \mathcal{I}_k \]

Attack Policy
In focus today

- Zero dynamics [S. Amin]
- Bias injection [A. Teixeira]
- DoS [M. Bishop]
- Eavesdropping [M. Bishop]
- Replay [B. Sinopoli]
- Covert [R. Smith]

Attack Space

Model knowledge

Disruption resources

Disclosure resources

K = {P, F, D}

a_k = g(K, I_k)

Disruption Resources

Disclosure Resources

[Teixeira et al., Automatica, 2015]
[Teixeira et al., HiCoNS, 2012]
Why Risk Management?

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 possible attack scenarios

What scenarios to prioritize?

What components to protect/defend first?
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 1: Power System State Estimator
Example 1: Power System State Estimator

Small security index $\alpha$ (to be defined) indicates sensors with inherent weak redundancy ($\sim$security). These should be defended first!

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

- Risk management
- **Attack detectability and security metric**
- Attack identification and secure state estimation
- Security metric computation
Basic Notions:  
Input Observability and Detectability

\[ x(k + 1) = Ax(k) + Bu(k), \quad x(k) \in \mathbb{R}^n, \quad u(k) \in \mathbb{R}^m \]
\[ y(k) = Cx(k) + Du(k), \quad y(k) \in \mathbb{R}^p \]

Definitions:

1. The input \( u \) is observable with knowledge of \( x(0) \) if \( y(k) = 0 \) for \( k \geq 0 \) implies \( u(k) = 0 \) for \( k \geq 0 \), provided \( x(0) = 0 \)

2. The input \( u \) is observable if \( y(k) = 0 \) for \( k \geq 0 \) implies \( u(k) = 0 \) for \( k \geq 0 \) (\( x(0) \) unknown)

3. The input \( u \) is detectable if \( y(k) = 0 \) for \( k \geq 0 \) implies \( u(k) \to 0 \) for \( k \to \infty \) (\( x(0) \) unknown)

[Hou and Patton, Automatica, 1998]
Basic Notions: Input Observability and Detectability

The Rosenbrock system matrix:

\[ P(z) = \begin{bmatrix} A - zI & B \\ C & D \end{bmatrix} \in \mathbb{C}^{(n+p) \times (n+m)} \]

First observations:

• Necessary condition for Definitions 1-3
  \[ \max_z \text{rank } P(z) = m + n \iff \text{normalrank } P(z) = m + n \]

• Fails if number of inputs larger than number of outputs \((m > p)\)

• Necessary and sufficient conditions involve the invariant zeros:
  \[ \sigma(P(z)) := \{ z : \text{rank } P(z) < \text{normalrank } P(z) \} \]
  (Transmission zeros + uncontrollable/unobservable modes, Matlab command: \texttt{tzero})
Basic Notions: Input Observability and Detectability

\[ P(z) = \begin{bmatrix} A - zI & B \\ C & D \end{bmatrix} \in \mathbb{C}^{(n+p) \times (n+m)} \]

**Theorems.** Suppose \((A, B, C, D)\) is minimal realization.

1. The input \(u\) is *observable with knowledge of* \(x(0) \iff \max_z \text{rank } P(z) = m + n \iff \text{normal rank } P(z) = m + n \)

2. The input \(u\) is *observable* \(\iff \forall z : \text{rank } P(z) = m + n \)
   (no invariant zeros)

3. The input \(u\) is *detectable* \(\iff (1) \) and
   \[ \sigma(P(z)) \subseteq \{z : |z| < 1\} \]
   (invariant zeros are all stable = system is minimum phase)

[Hou and Patton, Automatica, 1998]
Basic Notions: Input Observability and Detectability

\[ P(z) = \begin{bmatrix} A - zI & B \\ C & D \end{bmatrix}, \quad O(z) = \begin{bmatrix} A - zI \\ C \end{bmatrix} \]

**Theorems.** \((A, B, C, D)\) possibly non-minimal realization

1. The input \(u\) is observable with knowledge of \(x(0)\) \(\iff\)
   \[
   \max_z \text{rank} \ P(z) = m + n \iff \text{normal rank} \ P(z) = m + n
   \]

2’. The input \(u\) is observable \(\iff\) (1) and
   \[\sigma(P(z)) = \sigma(O(z))\]
   (invariant zeros are all unobservable modes)

3’. The input \(u\) is detectable \(\iff\) (1) and
   \[\sigma(P(z)) \setminus \sigma(O(z)) \subseteq \{z : |z| < 1\}\]
   (invariant zeros that are not unobservable modes are all stable)

[Hou and Patton, Automatica, 1998]
Fault Detection vs. Secure Control

Typical condition used in fault detection/fault tolerant control:

1. The input $u$ is observable with knowledge of $x(0) \Leftrightarrow$
   \[
   \max_z \text{rank } P(z) = m + n \Leftrightarrow \text{normal rank } P(z) = m + n
   \]
   \[\text{[Ding, Patton]}\]

Typical conditions used in secure control/estimation:

2. The input $u$ is observable $\Leftrightarrow$
   \[
   \forall z : \text{rank } P(z) = m + n
   \]
   (no invariant zeros)
   \[\text{[Sundaram, Tabuada]}\]

3/3’. The input $u$ is detectable $\Leftrightarrow$ (1) and
   \[
   \sigma(P(z)) \subseteq \{ z : |z| < 1 \}
   \]
   (invariant zeros are all stable = system is minimum phase)
   \[\text{[Pasqualetti, Sandberg]}\]
Example 2

\[ A = \begin{pmatrix} 0.9 & 0 & 0 \\ 0 & 0.8 & 0 \\ 0 & 0 & 0.9 \end{pmatrix}, \quad B = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \\ 0 & 0.25 \end{pmatrix}, \quad C = \begin{pmatrix} 0.4 & 0.6 & 0 \\ 0.2 & 0 & 0.4 \end{pmatrix} \]

\[ G(z) = C(zI - A)^{-1}B + D = \begin{pmatrix} \frac{0.2}{z-0.9} & \frac{0.3}{z-0.9} \\ \frac{0.1}{z-0.9} & \frac{0.1}{z-0.9} \end{pmatrix} \]

Invariant zeros = \( \sigma(P(z)) = \{1.1\} \)

[Note: normalrank \( P(z) = n + \) normalrank \( G(z) \)]

1. The input \( u \) is observable with knowledge of \( x(0) \): Yes!
2. The input \( u \) is observable: No!
3. The input \( u \) is detectable: No!

With \( x(0) = \begin{pmatrix} -0.705 \\ 0.470 \\ 0.352 \end{pmatrix} \) and \( u(k) = 1.1^k \begin{pmatrix} -0.282 \\ 0.282 \end{pmatrix} \) then \( y(k) = 0, k \geq 0 \)

OK for fault detection but perhaps not for security!
Attack and Disturbance Model

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

$$
x(k + 1) = Ax(k) + B_d d(k) + B_a a(k)
$$

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

- Unknown state $x(k) \in \mathbb{R}^n$ ($x(0)$ in particular)
- 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 $a(k) \not\to 0$ 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 $y(k) = 0$, $k \geq 0$ (Cf. Theorem 3')
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 \( a(k) = z_0^k a_0, 0 \neq a_0 \in \mathbb{C}^m, z_0 \in \mathbb{C} \), is undetectable iff there exists \( x_0 \in \mathbb{C}^n \) and \( d_0 \in \mathbb{C}^o \) such that

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

- Attack signal is undetectable if indistinguishable from measurable \((y)\) effects of natural noise \((d)\) or uncertain initial states \((x_0)\) [masking]
Example 2 (cont’d)

\[
G(z) = C(zI - A)^{-1}B + D = \begin{pmatrix} G_d(z) & G_a(z) \end{pmatrix}
\]

\[
G_d(z) = \begin{pmatrix} \frac{0.2}{z-0.9} & \frac{0.3}{z-0.8} \\ \frac{0.1}{z-0.9} & \frac{0.1}{z-0.9} \end{pmatrix}
\]

Poles = \{0.9, 0.9, 0.8\}

Invariant zeros = \( \sigma(P(z)) \) = \{1.1\}

Undetectable attack: \( a(k) = 1.1^k \begin{pmatrix} -0.282 \\ 0.282 \end{pmatrix} \)

Masking initial state: \( x_0 = \begin{pmatrix} -0.705 \\ 0.470 \\ 0.352 \end{pmatrix} \)
Example 3: Stealthy Water Tank Attack

2 hacked actuators \((u_1 \text{ and } u_2)\)

2 healthy sensors \((y_1 \text{ and } y_2)\)

Can the controller/detector always detect the attack?

[Teixeira et al., HiCoNS ‘12]
Example 3: Stealthy Water Tank Attack [Movie]
Example 3: Stealthy Water Tank Attack

2 hacked actuators ($u_1$ and $u_2$)
2 healthy sensors ($y_1$ and $y_2$)

Can the controller/detector always detect the attack?

Not against an adversary with physics knowledge
⇒ Undetectable attack exists (Similar to Example 2!)
Undetectable Attacks and Masking (cont’d)

• Suppose operator observes the output $y(k)$, and does not know the true initial state $x(0)$ and true disturbance $d(k)$
• Let $(x_0, d_0, a_0)$ be an undetectable attack, $0 = G_d d_0 + G_a a_0$ with initial state $x_0$

Consider the cases:
1. Un-attacked system $y = G_d (-d_0)$, with initial state $x(0) = 0$
2. Attacked system $y = G_a a_0$, with initial state $x(0) = x_0$

If initial states $x(0) = 0$ and $x(0) = x_0$ and disturbances $d = -d_0$ and $d = 0$ are equally likely, then impossible for operator to decide which case is true ⇒ Attack is undetectable!
Undetectable Attacks and Masking (cont’d)

• Suppose operator observes the output $y(k)$, and *knows* the true initial state $x(0) = 0$ and the disturbance $d(k) = 0, k \geq 0$
• Suppose system is asymptotically stable, $\rho(A) < 1$
• Let $(x_0, a_0)$ be an undetectable attack, $0 = G_a a_0$ with initial state $x_0$

Consider the cases:
1. Un-attacked system $y_1(k) = 0, k \geq 0$, with initial state $x(0) = 0$
2. Attacked system $y_2(k) = (G_a a_0)(k) = -CA^k x_0 \to 0$ as $k \to \infty$, with initial state $x(0) = 0$

The attacked output $y_2$ is vanishing, and can be made arbitrarily close to $y_1$ by scaling $(x_0, a_0) \Rightarrow \text{Attack is asymptotically undetectable!}$
The Security Index $\alpha_i$

$$\alpha_i := \min_{|z_0| \geq 1, x_0, d_0, a^i_0} \|a^i_0\|_0$$

subject to

$$P(z_0) \begin{bmatrix} x_0 \\ d_0 \\ a^i_0 \end{bmatrix} = 0$$

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

Interpretation:

- Attacker persistently targets signal component $a_i$ (condition $|z_0| \geq 1$)
- $\alpha_i$ is smallest number of attack signals that need to be simultaneously accessed to stage undetectable attack against signal $a_i$

Problem NP-hard in general (combinatorial optimization, cf. matrix spark). Generalization of static index in [Sandberg et al., SCS, 2010]
Example 4: Simple Security Index

\[ x(k + 1) = Ax(k) + B_d d(k) + B_a a(k) \]
\[ y(k) = C x(k) + D_d d(k) + D_a a(k) \]

\[
P(z) = \begin{bmatrix} A - zI & B_d & B_a \\ 0 & 0 & D_a \end{bmatrix} \quad D_a = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

- Measurements not affected by physical states and disturbances
- 3 measurements
- 4 attacks with security indices:
  - \( \alpha_1 = 3 \)
  - \( \alpha_2 = 3 \)
  - \( \alpha_3 = 3 \)
  - \( \alpha_4 = \infty \) (By definition. Even access to all attack signals not enough to hide attack)
Special Case 1: Critical Attack Signals

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

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

**Simple test, $\forall i$:** 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 ($o \geq p$)!
Special Case 2: Transmission Zeros

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

[\text{Amin et al., ACM HSCC, 2010}]
[\text{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 \( \alpha_i := \min_{|z_0| \geq 1, x_0, d_0, a_i^0} \|a_i^0\|_0 \)

subject to \( P(z_0) \begin{bmatrix} x_0 \\ d_0 \\ a_i^0 \end{bmatrix} = 0 \)

by inspection of corresponding zero directions \( \Rightarrow \text{Easy in typical case of 1-dimensional zero directions} \)
Example 2 (cont’d)

\[ G(z) = C(zI - A)^{-1}B + D = \begin{pmatrix} G_d(z) & G_a(z) \end{pmatrix} \]

\[ G_d(z) = (\begin{array}{c} \frac{0.2}{z-0.9} \\ \frac{0.1}{z-0.9} \end{array}) \quad G_a = \begin{pmatrix} \frac{0.3}{z-0.8} \\ \frac{0.1}{z-0.8} \end{pmatrix} \]

Invariant zeros = \( \sigma(P(z)) = \{1.1\} \)

Undetectable attack: \( a(k) = 1.1^k \begin{pmatrix} -0.282 \\ 0.282 \end{pmatrix} \Rightarrow a_0 = \begin{pmatrix} -0.282 \\ 0.282 \end{pmatrix} \)

Masking initial state: \( x_0 = \begin{pmatrix} -0.705 \\ 0.470 \\ 0.352 \end{pmatrix} \)

Only one signal satisfies \( \alpha_i \) constraint! Since \( ||a_0||_0 = 2 \Rightarrow \alpha_{1,2} = 2 \)
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** \( \alpha_i \)

**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 eigenvalues stable, no persistent undetectable sensor attack exists)
- Redundant measurements increase \( \alpha_i \)!
Special Case 4: Sensor Attacks for Static Systems

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

\[ (I - zI) = \begin{bmatrix} I_{n 	imes n} & 0 \\ C_{n 	imes \ell} & 0 \\ D_a_{\ell 	imes \ell} \end{bmatrix} \]

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

Space of eigenvectors \( x_0 \) is \( n \)-dimensional \( \Rightarrow \) Typically makes computation of \( \alpha_i \) 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 \( \alpha_i \) using min-cut/max-flow and \( \ell_1 \)-relaxation ([Hendrickx et al., 2014], [Kosut, 2014], [Yamaguchi et al., 2015])
Example 4: Simple Security Index

\[ P(\bar{z}) = \begin{bmatrix} A - zI & B_d & B_a \\ 0 & 0 & D_a \end{bmatrix} \]

\[ D_a = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \]

- Measurements not affected by physical states and disturbances
- 3 measurements
- 4 attacks with security indices:
  - \( \alpha_1 = 3 \)
  - \( \alpha_2 = 3 \)
  - \( \alpha_3 = 3 \)
  - \( \alpha_4 = \infty \) (By definition. Even access to all attack signals not enough to hide attack)
Special Case 4: Solution by MILP

Big $M$ reformulation:

\[
\alpha_i := \min_{x_0, a_0} \|a_0\|_0 \\
\text{subject to} \\
0 = Cx_0 + Da_0 \\
a_{0,i} = 1
\]

\[
\Rightarrow
\alpha_i := \min_{z_0, x_0, a_0, \sum k} z_k \\
\text{subject to} \\
0 = Cx_0 + Dxa_0 \\
a_{0,i} = 1 \\
-Mz \leq a_0 \leq Mz \\
z_k \in \{0, 1\}
\]

Elementwise

"∞"
Example 1: Power System State Estimator for IEEE 118-bus System

- State dimension \( n = 118 \)
- Number sensors \( p \approx 490 \)
Example 1: 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
- Used for defense allocation in [Vukovic et al., IEEE JSAC, 2012]
Summary So Far

• Basic risk management and control system attack space

• Dynamical security index $\alpha_i$ defined
  • Computation is NP-hard in general, but often “simple” in practically relevant cases:
    – One-dimensional zero-dynamics [Cases 2-3]
    – Static systems with special matrix structures (potential flow problems) [Case 4]
    – Dynamical models generally simplifies computation(!)
    – Redundant sensors increase $\alpha_i$

• Fast computation enables greedy security allocation
Outline

• Risk management

• Attack detectability and security metric

• Attack identification and secure state estimation

• Security metric computation
Attack Identification

\[ x(k + 1) = Ax(k) + B_d d(k) + B_a a(k) \]
\[ y(k) = C x(k) + D_d d(k) + D_a a(k) \]

- 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 \( a \) and \( \tilde{a} \) with \( \tilde{a}_i \neq a_i \), 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 \( \equiv \) (different attack \( a \) \( \Rightarrow \) different measurement \( y \)) \( \equiv \) attack signal is injectively mapped to \( y \) \( \Rightarrow \) attack signal is detectable

- **i-identifiable** weaker than **identifiable**

- \( \forall i: a \ is \ i\text{-identifiable} \iff a \ is \ identifiable \)

- \( a \ is \ i\text{-identifiable}: \) 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 [Sandberg and Teixeira, SoSCYPS, 2016] for details.
Example 4: Simple Security Index (cont’d)

\[ P(z) = \begin{bmatrix} A - zI & B_d & B_a \\ 0 & 0 & D_a \end{bmatrix} \quad D_a = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \]

Security indices: \( \alpha_1 = 3, \alpha_2 = 3, \alpha_3 = 3, \alpha_4 = \infty \)

Attacker with \( q = 1 \): Defender can identify (and thus detect) all attacks

\( q = 2 \): Defender can detect (not identify) all attacks against \( a_1, a_2, a_3 \) and identify all attacks against \( a_4 \)

\( q = 3 - 4 \): Defender can identify all attacks against \( a_4 \). Exist undetectable attacks against \( a_1, a_2, a_3 \)
Back to Risk Management

\[
P(z) = \begin{bmatrix} A - zI & B_d & B_a \\ 0 & 0 & D_a \end{bmatrix} \quad D_a = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

Security indices: \( \alpha_1 = 3, \quad \alpha_2 = 3, \quad \alpha_3 = 3, \quad \alpha_4 = \infty \)

- Suppose the operator can choose to block one attack signal (through installing physical protection, authentication, etc.).
- Which signal \( a_1, a_2, a_3, \) or \( a_4 \) should she/he choose?
- Among the one(s) with lowest security index! Choose \( a_1 \).
- New attack model and security indices: \( \alpha_2 = \alpha_3 = \alpha_4 = \infty \)

\[
D_a = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

- By explicitly blocking one attack signal, all other attacks are implicitly blocked (they are identifiable)
Example 1: 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 $y$
Outline

• Risk management

• Attack detectability and security metric

• Attack identification and secure state estimation

• Security metric computation
DC-Power Flow Measurement Matrix

\[ C = \begin{bmatrix} P_1 DB^T \\ -P_2 DB^T \\ P_3 BDB^T \end{bmatrix} \]

(positive flow measurements)

(negative flow measurements)

(injection measurements)

\( B \)- directed incidence matrix of graph corresponding to power network topology

\( D \)- nonsingular diagonal matrix containing reciprocals of reactance of transmission lines

\( P_\cdot \)- measurement selection matrices (rows of identity matrices)

More measurements than states, \( m > n \). Redundancy!

Structure applies to all potential flow problems (water, gas, …)

[Hendrickx et al., IEEE TAC, 2014]
Example 5: DC-Power Flow Measurement Matrix

\[
B^T = \begin{pmatrix}
1 & -1 & 0 \\
1 & 0 & -1 \\
0 & 1 & -1
\end{pmatrix}, \quad D = I
\]

Edge flows: \[
\begin{pmatrix}
p_{12} \\
p_{13} \\
p_{23}
\end{pmatrix} = DB^T x = \begin{pmatrix}
x_1 - x_2 \\
x_1 - x_3 \\
x_2 - x_3
\end{pmatrix}
\]

Node injections: \[
BDB^T x = B \begin{pmatrix}
p_{12} \\
p_{13} \\
p_{23}
\end{pmatrix} = \begin{pmatrix}
p_{12} + p_{13} \\
-p_{12} + p_{23} \\
-p_{13} - p_{23}
\end{pmatrix}
\]
Example 5: DC-Power Flow Measurement Matrix (cont’d)

\[ C = \begin{bmatrix} P_1 DB^T \\ -P_2 DB^T \\ P_3 BDB^T \end{bmatrix} \]

\[ B^T = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \quad D = I \]

\[
P_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

\[
P_2 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}
\]

\[
P_3 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}
\]
Efficient Security Index Computation

Rewrite security index problem into equivalent form:

\[ J_c := \min_{x \in \mathbb{R}^n} c^T g(DB^T x) + p^T g(BDB^T x) \equiv \|Cx\|_0 \]

subject to \( B(:,k)^T x \neq 0 \) (enforce non-zero flow on edge \( k \))

- \( g(\cdot) \) - Vector-valued indicator function (ex. \( g \begin{pmatrix} -3 \\ 0 \\ 1.5 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \))
- \( c \in \mathbb{R}_+^r \) - Encodes #edge flow meters (\( r \) edges)
- \( p \in \mathbb{R}_+^n \) - Encodes #node injection meters (\( n \) nodes/states)

Choose index \( k \) to activate sensor \( i \) so that \( J_c = \alpha_i \)
Restricted Binary Problem

\[
J_c := \min_{x \in \mathbb{R}^n} c^T g(DB^T x) + p^T g(BDB^T x) \equiv \|Cx\|_0
\]

subject to \(B(:, k)^T x \neq 0\) (enforce non-zero flow on edge \(k\))

\[
J_b := \min_{x \in \{0,1\}^n} c^T g(DB^T x) + p^T g(BDB^T x) \equiv \|Cx\|_0
\]

subject to \(B(:, k)^T x \neq 0\) (enforce non-zero flow on edge \(k\))

Obviously \(J_b \geq J_c\), but in fact we have...

**Theorem.**

\[
0 \leq J_b - J_c \leq \sum_{i=1}^{n} \max\{0, \max_{e \to v_i} \{p_i - c_e\}\}
\]

\((e \to v_i := \text{set of edges connected to node with state } x_i.\)\)

[Hendrickx et al., IEEE TAC, 2014]
Example 5: DC-Power Flow Measurement Matrix (cont’d)

\[ c = \begin{pmatrix} 1 & 1 & 2 \\ \end{pmatrix} \]

\[ p = \begin{pmatrix} 1 & 0 & 0 \\ \end{pmatrix} \]

\[ 0 \leq J_b - J_c \leq \sum_{i=1}^{n} \max \left\{ 0, \max_{e \to v_i} \left\{ p_i - c_e \right\} \right\} \]

\[ = \max \{0, \max \{0, 0\}\} + \max \{0, \max \{-1, -2\}\} + \max \{0, \max \{-1, -2\}\} \]

\[ = 0 \]

Solve \( J_b \) to find exact security indices \( \alpha_i \)!
Example 5: DC-Power Flow Measurement Matrix (cont’d)

\[ c = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix} \]
\[ p = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \]

Compute \( \alpha_1 \) in 4 steps:

1. Enforce flow across sensor 1 by choosing \( x_1 = 1 \) and \( x_2 = 0 \) \([J_b \text{ constraint satisfied}]\)
2. Test \( x_3 = 0 \): \( ||Cx||_0 = 3 \)
3. Test \( x_3 = 1 \): \( ||Cx||_0 = 4 \)
4. \( J_b = \min\{3,4\} = 3 = \alpha_1 \)
Summary

• There is a need for CPS security
• Briefly introduced CPS attack models and concept of risk management

• Input observability and detectability
  ⇒ Undetectable attacks and masking initial states and disturbances

• A security metric $\alpha_i$ for risk management
  • Suppose attacker has access to $q$ resources:
    – Undetectable attacks against $a_i$ iff $q \geq \alpha_i$
    – Attack against $a_i$ identifiable iff $q < \alpha_i/2$

• Many useful results in the fault diagnosis literature, especially for identifiable attacks: Unknown input observers, decoupling filters, etc.
• Future research direction: More realistic attacker models, estimate attack likelihoods and impacts, corporation with IT security,…
Further Reading

Introduction to CPS/NCS security

• Special Issue on CPS Security, IEEE Control Systems Magazine, February 2015

CPS attack models, impact, and risk management

Further Reading

Detectability and identifiability of attacks


Further Reading

Security metrics (security index)


Acknowledgments

André M.H. Teixeira
(Delft University of Technology)

Kin Cheong Sou
(National Sun Yat-sen University)

György Dán
Karl Henrik Johansson
Jezdimir Milošević
David Umsonst
(KTH)
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:

$$
\begin{bmatrix}
B_d \\
D_d
\end{bmatrix} d + \begin{bmatrix}
B_a \\
D_a
\end{bmatrix} a = \begin{bmatrix}
B_f \\
D_f
\end{bmatrix} f,
$$

such that $\begin{bmatrix}
B_f \\
D_f
\end{bmatrix}$ full column rank

2. Compute security indices $\alpha_i$ 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 

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

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 \( \Delta_i \) of \( \Delta \), for all subsets \( |I| = q, I \subseteq \{1, \ldots, m\} \) (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)