Secure control systems: An overview

André Teixeira
Uppsala University, Sweden
andre.teixeira@angstrom.uu.se
www.andre-teixeira.eu

Plenary Talk, SysTol’19, Sep. 19
Outline

• Motivation

• Risk management
  • Scenario characterisation
  • Risk Analysis
  • Risk Mitigation

• Secure control: from analysis to design
  • Problem formulation
  • Metrics in Fault-Tolerant Control
  • Metrics in Secure Control
    • Computation and design problem
Cyber-Physical Systems

Industrial Control System (ICS)

Cloud-based Control and IoT

Autonomous vehicles
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

• Zero-day vulnerabilities
Example 1: Stuxnet (2010)


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.
Example 2: Triton Malware (2017)

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.
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, and how?
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
- **Can we trust** the interfaces and channels are really secured? (see OpenSSL Heartbleed bug…)
- **Malicious actions can enter anywhere** in the closed loop and cause harm
- **Malicious attackers** have an **intent**, as opposed to faults, and can act strategically
• Motivation

• Risk management
  • Scenario characterisation
  • Risk Analysis
  • Risk Mitigation

• Secure control: from analysis to design
  • Problem formulation
  • Metrics in Fault-Tolerant Control
  • Metrics in Secure Control
    • Computation and design problem
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, and how?
- How to find answers: 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 (or ~1 / complexity of attack)

• **Impact**
  – What are the cyber-physical consequences of an attack?

[Kaplan & Garrick, 1981], [Bishop, 2002]
Main steps in risk management

- Scenario characterization
  - Models, Scenarios, Objectives

- Risk Analysis
  - Likelihood Assessment
  - Impact Assessment

- Risk Mitigation
  - Prevention, Detection, Treatment
Networked Control System under Attack

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

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

[Teixeira et al., Automatica, 2015]
Adversary Model

- **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]
Attack Space

Disruption resources

Undetectable attack

CPS model knowledge

Covert attack

[22]–[24]

[24], [25]

Bias injection attack

DoS attack

[26]–[32]

[15], [33]–[47]

Eavesdropping attack

[48]–[50]

Replay attack

[Chong et al., ECC, 2019]

Model Knowledge

$\mathcal{K} = \{\hat{P}, \hat{F}, \hat{D}\}$

$\alpha_k = g(\mathcal{K}, \mathcal{I}_k)$

Disclosure Resources

$\mathcal{I}_k \Rightarrow y_k$

Disruption Resources

$\hat{P} \Rightarrow \hat{F} \Rightarrow \hat{D}$

Attack Policy

Secure Control Systems: an overview

André Teixeira, Sep. 19, 2019
Example: Undetectable Water

2 hacked actuators ($u_1$ and $u_2$ = disruption resources)

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

Can the controller/detector always detect the attack?

[Teixeira et al., Automatica, 2015]
Undetectable Water Tank Attack [Movie]

Secure Control Systems: an overview

André Teixeira, Sep. 19, 2019
Water Tank Model Analysis

- Transfer function matrix from attack to sensor signals
  \[ G_a(z) = C(zI - A)^{-1}B = \begin{pmatrix} \frac{0.0289}{z-0.8076} & \frac{(1.277z+1.182)\cdot10^{-3}}{z^2-1.784z+0.7928} \\ \frac{(1.356z+1.24)\cdot10^{-3}}{z^2-1.754z+0.7643} & \frac{0.02954}{z-0.8347} \end{pmatrix} \]

- Poles = \{0.8076, 0.8347, 0.9464, 0.9498\}
- Invariant zeros = \{0.8675, 1.0362\} \Rightarrow \text{Non-minimum phase system}
- Applied attack signal (small \(\epsilon\))
  \[ a(k) = 1.0362^k \begin{pmatrix} 0.2281\epsilon \\ -0.2281\epsilon \end{pmatrix}, \quad x_0 = \begin{pmatrix} 0 \\ 0 \\ -0.6521\epsilon \\ 0.6876\epsilon \end{pmatrix}^T \]
  satisfies \textbf{zero dynamics} and is \textbf{masked by} system transient:
  \[ 0 = y(k) = CA^kx_0 + (g_a * a)(k), \quad k \geq 0 \]
Undetectable Water Tank Attack

2 hacked actuators ($u_1$ and $u_2$ = disruption resources)

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

Can the controller/detector always detect the attack?

Not against an adversary with full CPS model knowledge

[Teixeira et al., Automatica, 2015]
The Rosenbrock system matrix: 

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

**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, TAC, 2013], [Sandberg and Teixeira, SoSCYPS, 2016]
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

\[ \alpha_i := \min_{|z_0| \geq 1, x_0, d_0, a'_0} \|a'_0 \|_0 \]

subject to \[ P(z_0) \begin{bmatrix} x_0 \\ d_0 \\ a'_0 \end{bmatrix} = 0 \]

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 an undetectable attack against component \( a_i \)
- Estimate likelihood for attack against component \( i \) by \( \sim 1/\alpha_i \)
- Problem NP-hard, but easy when geometric multiplicities of zeros are 1

[Sandberg and Teixeira, SoSCYPS, 2016]
Security Index: Water Tank Example

\[ G_a(z) = C(zI - A)^{-1}B = \begin{pmatrix} \frac{0.0289}{z-0.8076} & \frac{(1.277z+1.182)\cdot10^{-3}}{z^2-1.784z+0.7928} \\ \frac{(1.356z+1.24)\cdot10^{-3}}{z^2-1.754z+0.7643} & \frac{0.02954}{z-0.8347} \end{pmatrix} \]

- Invariant zeros = \{0.8675, 1.0362\} \Rightarrow \text{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 \( \alpha_i \) constraint!
  \[ \|a(k)\|_0 = 2 \Rightarrow \alpha_{1,2} = 2 \]
Security Index
IEEE 14 Bus System

[Milosevic et al., arXiv preprint, 2019]
System model:

\[ x(k + 1) = Ax(k) + B_a a(k) \]
\[ y(k) = Cx(k) + D_a a(k) \]

Consider trajectories from \( k=0, \ldots, N \).

Impact assessment problem:

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

Theorem 3: \[\text{[Milosevic et al., ECC, 2018]}\]

i. Constraints are convex

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

• **Prevention** (decrease likelihood by reducing vulnerability)
  - Watermarking and Moving Target Defense
  - Coding and Encryption Strategies
  - Rational Security Allocation
  - Privacy-preservation by Noise Injection

• **Detection** (continuous anomaly monitoring)
  - Tuning of Detector Thresholds
  - Secure State Estimation
  - Watermarking and Moving Target Defense
  - Distributed Algorithms
  - Methods Related to Robust Statistics

• **Treatment** (compensate for or neutralize detected attack)
  - Secure State Estimation
  - Countering DoS Attacks
  - Distributed Algorithms
  - Methods Related to Robust Statistics

[Chong et al., ECC, 2019]
Outline

- Motivation

- Risk management
  - Scenario characterisation
  - Risk Analysis
  - Risk Mitigation

- Secure control: from analysis to design
  - Problem formulation
  - Metrics in Fault-Tolerant Control
  - Metrics in Secure Control
    - Computation and design problem
System model:

\[ x(k + 1) = Ax(k) + B_a a(k) \]
\[ y(k) = Cx(k) + D_a a(k) \]

Consider trajectories from \( k=0, \ldots, N \).

**Impact assessment problem:**

maximize \( a \) \( \|x\|_\infty \)

subject to \( a \in \{ \text{DoS, Data Injection, Re-routing, Replay, Bias} \} \) attack

\( y \) generates no alarm in \( \{ \chi^2, \text{CUSUM, MEWMA} \} \) detector

**Drawbacks:**

Only looks at the terminal state
It only serves for analysis, not for design…
Networked Control System under Attack

- Physical plant $\mathcal{P}$
- Feedback controller $\mathcal{F}$
- Anomaly detector $\mathcal{D}$

- Performance output: $y_p(t)$ (e.g. physical state)
- Detector output: $y_r(t)$

- Actuator and Sensor data corruption: $a(t) = \begin{bmatrix} \Delta u(t) \\ \Delta y_m(t) \end{bmatrix}$

\[
\|y_r\|_2^2 \geq 1 \Rightarrow \text{Alarm!}
\]

Closed-loop system

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Ba(t) \\
y(t) &= Cx(t) + Da(t)
\end{align*}
\]
Adversary Model

- **Model knowledge**: Dynamical model of the closed-loop system
- **Disruption resources**: (Small no. of) measurement and actuation channels
- **Attack policy**: Maximise the impact on performance without raising alarms

How to analyse and design the system with such an attack?
Fault-Tolerant Control
Fault-Tolerant Control

- While fault is not detected:
  - Robust controller

- Once fault is detected:
  - Switch/adapt controller

FTC

- Passive
  - Robust control

- Active
  - FDI/System Identification
  - Control reconfiguration/Restructure
  - Projection
  - Online controller redesign/adaptation
FTC design objectives

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

• **Fault detection**: find an observer/filter that
  • Maximizes the “worst-case” (smallest) **detectability** of unit-energy faults
  • i.e.: optimal “H minus” (H\(_{-}\)) filter design

• Both are based on **sensitivity metrics**:
  • H-inf norm: largest **impact** of unit-energy faults
  • H\(_{-}\) index: smallest **detectability** of unit-energy faults
Classical Sensitivity Metrics (1)

\( L_2 = \text{“signals with finite energy over finite time intervals”} \)

\[ \|y\|_{L_2}^2 \triangleq \int_{-\infty}^{+\infty} \|y(t)\|_2^2 \, dt \]

- **H-inf norm:**
  \[ \gamma_{H_\infty} \triangleq \sup_{a \in \mathcal{L}_2} \|y_p\|_{L_2} \]
  s.t. \( \|a\|_{L_2} = 1 \)
  \( x(0) = 0 \)

- **Frequency Domain:**
  \[ \gamma_{H_\infty} = \sup_{w \geq 0} \bar{\sigma}_p(jw) \]
  \[ \bar{\sigma}_p(s) = \sup_{a \in \mathbb{C}^n} \|G_p(s)a\|_2 \]
  s.t. \( \|a\|_2 = 1 \)

- **H_index:**
  \[ \gamma_{H_-} \triangleq \inf_{a \in \mathcal{L}_2} \|y_r\|_{L_2} \]
  s.t. \( \|a\|_{L_2} = 1 \)
  \( x(0) = 0 \)

- **Frequency Domain:**
  \[ \gamma_{H_-} = \inf_{w \geq 0} \sigma_r(jw) \]
  \[ \sigma_r(s) = \inf_{a \in \mathbb{C}^n} \|G_r(s)a\|_2 \]
  s.t. \( \|a\|_2 = 1 \)
Example: closed-loop system

- **Process**: Quadruple tank
  - (non-minimum phase setup, one unstable zeros from u to y)

- **Controller**: LQG with integral action
  - Performance output = plant’s states
    \[ y_p(t) = x_p(t) \]

- **Detector**: LQG’s Kalman filter
  - Residual = output estimation error
    \[ y_r(t) = y(t) - \hat{y}(t) \]

- The adversary is able to corrupt Actuator 1 (u₁)
Classical Sensitivity Metrics (2)

- **H-inf norm**
  \[ \gamma_{H_\infty} = \sup_{w \geq 0} \bar{\sigma}_p(jw) \]

- **H_- index**
  \[ \gamma_{H_-} = \inf_{w \geq 0} \sigma_r(jw) \]

- Least detectable fault has little impact…

- **Limitation:** worst-case frequency is not the same
  - Means each metric looks at **different** worst-case inputs!
Sensitivity metric for security of control systems (1)

- **Attack policy**: Maximise the impact on performance without raising alarms
  - Requires a combination of impact on performance and detection!

- “Desired” metric: combine H-inf and H_: \( \tilde{\gamma}^* = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) = \frac{\tilde{\sigma}_p(jw)}{\tilde{\sigma}_r(jw)} \)

- But would it be meaningful / accurate?

- And how do we formalise such a metric?

> Coming next!
Sensitivity metric for security of control systems (1)

- **Attack policy:** Maximise the impact on performance without raising alarms.
  - Requires a combination of impact on performance and detection!
- “Desired” metric: combine $H\text{-inf}$ and $H\_:$
  \[ \tilde{\gamma}^* = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\sigma_p(jw)}{\sigma_r(jw)} \]

- But this is an heuristic... When would it actually be meaningful / accurate?
- And how do we formalise such a metric?
  - ... coming next!
Sensitivity metric for security of control systems (2)

- **Attack policy:** Maximise the impact on performance without raising alarms
- Maximize $y_p$, while keeping $y_r$ small — **Output-to-output gain**:

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

[Teixeira et al., CDC 15]

- An equivalent formulation:

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

- Note: Input is not directly constrained (may be exponentially increasing)\n\[\mathcal{L}_{2e} = \text{“signals with finite energy over finite time intervals”}\n
- Imposes a gain constraint on the outputs:

\[
\gamma^*^2 \|y_r\|^2_{\mathcal{L}_2} \geq \|y_p\|^2_{\mathcal{L}_2}
\]
Computation: Linear Matrix Inequality

\[
\gamma^* = \min_{\beta \geq 0, P \geq 0} \beta \\
\text{s.t. } \quad \begin{bmatrix}
A^T P + PA & PB \\
B^T P & 0
\end{bmatrix} - \beta \begin{bmatrix}
C_r^T C_r & C_r^T D_r \\
D_r^T C_r & D_r^T D_r
\end{bmatrix} + \begin{bmatrix}
C_p^T C_p & C_p^T D_p \\
D_p^T C_p & D_p^T D_p
\end{bmatrix} \leq 0
\]

• The value of the metric can be obtained by solving this convex problem

• Controller / Detector design is possible
  • A, B, C, D depend linearly on the controller / detector
  • … but leads to Bilinear Matrix Inequalities
  • Exploit structure to convexify the problem - ongoing work

• Other metrics can be formulated in a similar fashion [Teixeira, CDC 19]
Summary

• Secure Control systems - a risk management approach

• Classical sensitivity metrics are not adequate for security-related problems / malicious adversary models
  • They focus either on impact or on detection separately

• Output-to-output gain captures both impact and detection
  • Maximize energy of ”cost signal”; While keeping ”anomaly detector signal” small
  • Computation based on LMIs
  • Controller / Detector design through BMIs

andre.teixeira@angstrom.uu.se

Thank you!

Secure Control Systems: an overview

André Teixeira, Sep. 19, 2019
References


Secure Control Systems: an overview

André Teixeira, Sep. 19, 2019
Extra slides
Example: Attack scenarios

- The adversary is able to corrupt:
  A. Actuator 1
  B. Actuator 1 and Actuator 2 (unstable zero $\rightarrow y=0$)
  C. Actuator 1 and Sensor 1
  D. Actuator 1 and Sensor 2

- Evaluate which scenarios are most critical (largest sensitivity)
\[
\gamma(s) \triangleq \sup_{a \in \mathbb{C}^{n_a}} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2} \quad \tilde{\gamma}^* = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\bar{\sigma}_p(jw)}{\bar{\sigma}_r(jw)}
\]

\[\gamma^* = \sup_{s \in \mathcal{S}} \gamma(s)\]

Example: Actuator 1

- Gain vs heuristic: \( \gamma^* = 23.8, \tilde{\gamma}^* = 23.8 \) (LMI gives: \( \gamma^* = 24.1 \))
- Gain over Frequency: \( \gamma(jw) = \tilde{\gamma}(jw) \)

\[
\sup_{w \geq 0} \gamma(jw) = 23.8
\]
\[ \gamma(s) \triangleq \sup_{a \in \mathbb{C}^n} \frac{\|G_p(s)a\|_2}{\|G_r(s)a\|_2} \]

\[ \gamma^* = \sup_{s \in S} \gamma(s) \]

\[ \tilde{\gamma}^* = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\tilde{\sigma}_p(jw)}{\tilde{\sigma}_r(jw)} \]

Example: Actuator 1 and Actuator 2

- **Gain vs heuristic:** \( \gamma^* = \infty, \tilde{\gamma}^* = 58.3 \) (unstable zero)
- **Gain over Frequency:** \( \gamma(jw) \leq \tilde{\gamma}(jw) \)

\[ \sup_{w \geq 0} \gamma(jw) = \tilde{\gamma}^* \]
Secure Control Systems: an overview  André Teixeira, Sep. 19, 2019

\[ \gamma(s) \triangleq \sup_{a \in \mathbb{C}^{n_a}} \frac{\| G_p(s)a \|_2}{\| G_r(s)a \|_2} \]

\[ \gamma^* = \sup_{s \in \mathcal{S}} \gamma(s) \]

\[ \tilde{\gamma}^* = \sup_{w \geq 0} \tilde{\gamma}(jw), \quad \tilde{\gamma}(jw) \triangleq \frac{\tilde{\sigma}_p(jw)}{\tilde{\sigma}_r(jw)} \]

Example: Actuator 1 and Sensor 1

- Gain vs heuristic:
  \[ \gamma^* = \infty, \quad \tilde{\gamma}^* = \infty \]

- Gain over Frequency:
  \[ \gamma(jw) \leq \tilde{\gamma}(jw) \]
  \[ \sup_{w \geq 0} \gamma(jw) = \gamma^* \] (zeros at infinity)
Example: Actuator 1 and Sensor 2

- Gain vs heuristic: \( \gamma^* = 73.1, \tilde{\gamma}^* = 253.5 \)
- Gain over Frequency: \( \gamma(jw) \leq \tilde{\gamma}(jw) \) \[ \sup_{w \geq 0} \gamma(jw) = \gamma^* \] (zeros at infinity are cancelled)