Attack Detection and Identification in Cyber-Physical Systems

Francesco Bullo

Center for Control,
Dynamical Systems & Computation
University of California at Santa Barbara
http://motion.me.ucsb.edu

International Workshop on Emerging Frontiers in Systems and Control
Tsinghua University, Beijing, China, May 18, 2012

Acknowledgements

Florian Dörfler

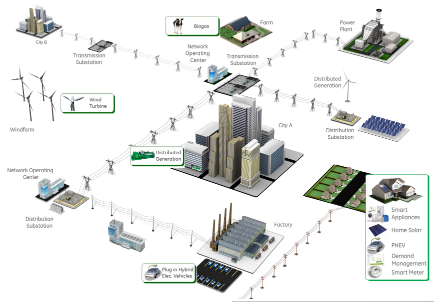
Fabio Pasqualetti

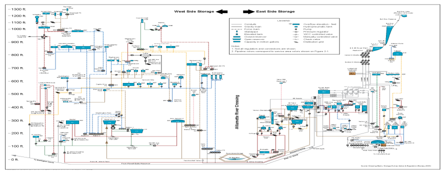
Workshop Organizers:
Center for Intelligent and Networked Systems, Tsinghua University
Institute of Systems Science, Chinese Academy of Sciences

Chair: Xiaohong Guan, Co-Chair: Yiguang Hong, Program Chair: Qingshan Jia

Outline

Cyber-Physical Systems

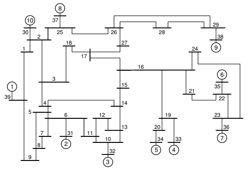


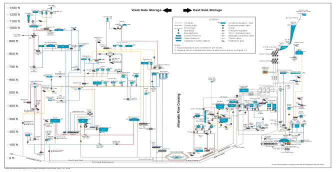


Moore's Law in Computing/Communication/Control

Renewables and PMUs in smart grid, autonomy/networking in robotics, distributed intelligence in industrial processes $\rightsquigarrow$ cyber-physical networks

Application Domains





- power generation, transportation, distribution networks
- water, oil, gas and mass transportation systems
- sensor networks
- process control and industrial automation systems (metallurgical process plants, oil refining, chemical plants, pharmaceutical manufacturing ... ubiquitous SCADA/PLC systems)

Security of these networks is critically important

The Cyber-Physical Security Problem

Stuxnet worm (Iran, 2010)

New York Times 15jan2011: replay attack as if "out of the movies:"

1. records normal operations and plays them back to operators
2. spins centrifuges at damaging speeds

"Repository of Ind. Security Incidents"
http://www.securityincidents.org

SOME OF MANY

Water industry
Maroochy Shire sewage spill; Salt River Project SCADA hack; software flaw makes MA water undrinkable; Trojan/Keylogger on Ontario SCADA System; viruses on Aussie SCADA laptops; audit/blaster causes water SCADA crash; penetration of California irrigation district wastewater treatment plant SCADA; SCADA system tagged with message: 'I enter in your server like you in Iraq'.

Petroleum industry
Electronic sabotage of Venezuela oil operations; CIA Trojan causes Siberian gas explosion; anti-virus software prevents boiler safety shutdown; slammer infected laptop shuts down DCS; electronic sabotage of gas processing plant; Slammer impacts offshore platforms; Code Red Worm defaces automation Web pages; penetration test locks-up gas SCADA System.

Chemical industry
IP address change shuts down chemical plant; hacker changes chemical plant set points; Nachi Worm on advanced process control servers; SCADA attack on plant of chemical company; contractor connects to remote PLC; Blaster Worm infects chemical plant.

Power industry
Slammer infects control central LAN via VPN; Slammer causes loss of comms to substations; Slammer infects Ohio nuclear plant SPDS; Iranian hackers attempt to disrupt Israel power system; utility SCADA System attacked; virus attacks a European Utility; facility cyber attacks on Asian utility; power plant security details leaked on Internet.

Cyber-Physical Security $\neq$ Cyber Security, Fault Tolerance

Cyber-physical security complements cyber security

Cyber security (e.g., secure communication, secure code execution)

- does not verify "data compatible with physics/dynamics"
- is ineffective against direct attacks on the physics/dynamics
- is never foolproof (e.g., insider attacks, OS zero-day vulnerabilities)

Cyber-physical security extends fault tolerance

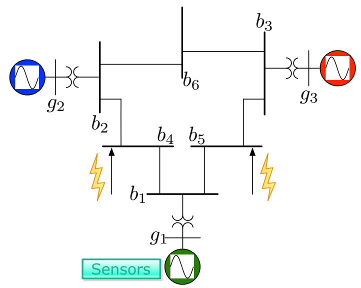
- fault detection considers *accidental/generic failures*
- cyber-physical security models *worst-case attacks*

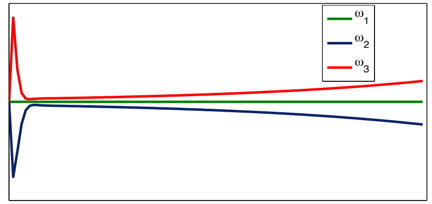
An Incomplete List of Related Results

- S. Amin et al, "Safe and secure networked control systems under denial-of-service attacks," *Hybrid Systems: Computation and Control* 2009.
- Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," *ACM Conference on Computer and Communications Security*, Nov. 2009.
- A. Teixeira et al. "Cyber security analysis of state estimators in electric power systems," *IEEE Conf. on Decision and Control*, Dec. 2010.
- S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Stealthy deception attacks on water SCADA systems," *Hybrid Systems: Computation and Control*, 2010.
- Y. Mo and B. Sinopoli, "Secure control against replay attacks," *Allerton Conf. on Communications, Control and Computing*, Sep. 2010
- G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," *IEEE Int. Conf. on Smart Grid Communications*, Oct. 2010.
- Y. Mo and B. Sinopoli, "False data injection attacks in control systems," *First Workshop on Secure Control Systems*, Apr. 2010.
- S. Sundaram and C. Hadjicostis, "Distributed function calculation via linear iterative strategies in the presence of malicious agents," *IEEE Transactions on Automatic Control*, vol. 56, no. 7, pp. 1495–1508, 2011.
- R. Smith, "A decoupled feedback structure for covertly appropriating network control systems," *IFAC World Congress*, Aug. 2011.
- F. Hamza, P. Tabuada, and S. Diggavi, "Secure state-estimation for dynamical systems under active adversaries," *Allerton Conf. on Communications, Control and Computing*, Sep. 2011.

Outline

An Example of Cyber-Physical Attack





1. **Physical dynamics:** classical generator model & DC load flow
2. **Measurements:** angle and frequency of generator g_1
3. **Attack:** modify real power injections at buses b_4 & b_5

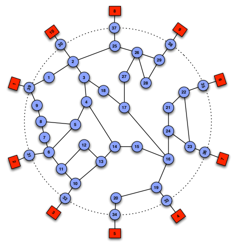
"Distributed internet-based load altering attacks against smart power grids" *IEEE Trans on Smart Grid, 2011*

The attack affects the second and third generators while remaining undetected from measurements at the first generator

Models of Power Networks

Small-signal structure-preserving power network model:

1. transmission network: generators ■, buses ●, DC load flow assumptions, and network susceptance matrix $Y = Y^T$

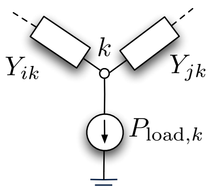


2. generators ■ modeled by swing equations:

$$M_i\ddot{\theta}_i + D_i\dot{\theta}_i = P_{\text{mech.in},i} - \sum\nolimits_j Y_{ij}\cdot(\theta_i - \theta_j)$$

3. buses ● with constant real power demand:

$$0 = P_{\text{load},i} - \sum\nolimits_j Y_{ij}\cdot(\theta_i - \theta_j)$$

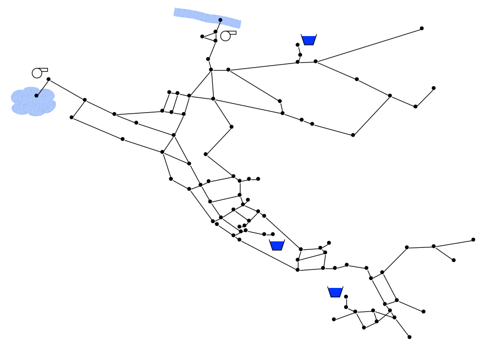


⇒ Linear differential-algebraic dynamics: $E\dot{x} = Ax$

Models of Water Networks

Linearized municipal water supply network model:

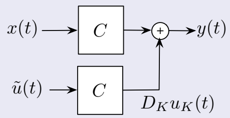
1. reservoirs with constant pressure heads: $h_i(t) = h_i^{\text{reservoir}} = const.$
2. pipe flows obey linearized Hazen-Williams eq: $Q_{ij} = g_{ij}\cdot(h_i - h_j)$
3. balance at tank: $A_i\dot{h}_i = \sum_{j\to i} Q_{ji} - \sum_{i\to k} Q_{ik}$
4. demand = balance at junction: $d_i = \sum_{j\to i} Q_{ji} - \sum_{i\to k} Q_{ik}$
5. pumps & valves: $h_j - h_i = +\Delta h_{ij}^{\text{pump/valves}} = \text{const.}$



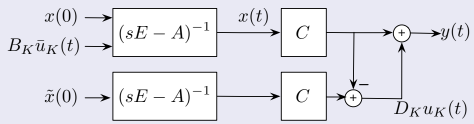
⇒ Linear differential-algebraic dynamics: $E\dot{x} = Ax$

Prototypical Attacks

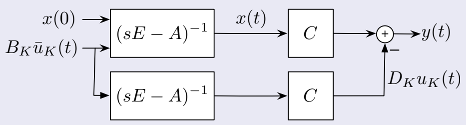
Static stealth attack:
corrupt measurements according to C



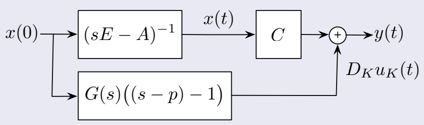
Replay attack:
affect system and reset output



Covert attack:
closed loop replay attack



Dynamic false data injection:
render unstable pole unobservable



Models of Networks, Attackers and Monitors #1

Network model

$$E\dot{x}(t) = Ax(t) + Bu(t) \qquad \text{(state and actuator attack)}$$

$$y(t) = Cx(t) + Du(t) \qquad \text{(data substitution attack)}$$

Byzantine Cyber-Physical Attackers

1. colluding omniscent attackers:
   - know model structure and parameters
   - measure full state
   - can apply some control signal and corrupt some measurements
2. attacker's objective is to change/disrupt the physical state

Models of Networks, Attackers and Monitors #2

Security System

1. knows structure and parameters
2. measures output signal

Objectives

1. vulnerability analysis (fundamental monitor limitations)
2. detection and identification monitors
3. secure-by-design systems
4. attack strategies

Outline

Framework for Cyber-Physical Security

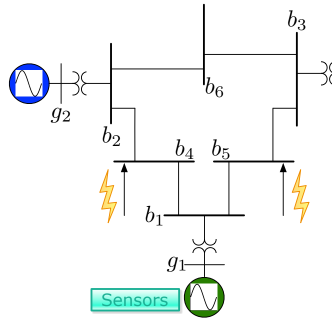
1. a modeling framework for cyber-physical systems under attack generalizing broad range of previous results
2. fundamental detection and identification limitations
3. system- and graph-theoretic detection and identification conditions
4. centralized attack detection and identification procedures
5. distributed attack detection and identification procedures

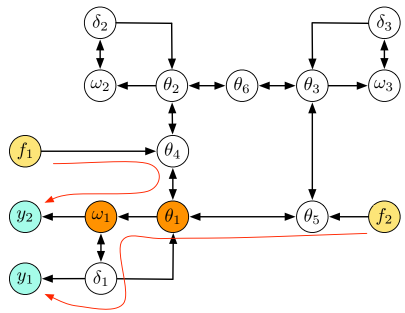
References

- F. Pasqualetti, F. Dorfler, and F. Bullo. "Cyber-physical security via geometric control: Distributed monitoring and malicious attacks" 2012 IEEE CDC. Submitted
- —— "Attack Detection and Identification in Cyber-Physical Systems – Part I: Models and Fundamental Limitations" IEEE Trans Automatic Control, Feb 2012. Submitted. Available at http://arxiv.org/abs/1202.6144v2
- —— "Attack Detection and Identification in Cyber-Physical Systems – Part II: Centralized and Distributed Monitor Design" IEEE Trans Automatic Control, Feb 2012. Submitted. Available at http://arxiv.org/abs/1202.6049

Result #1: Vulnerabilities Analysis

Western US (WECC 3-m, 6-b)

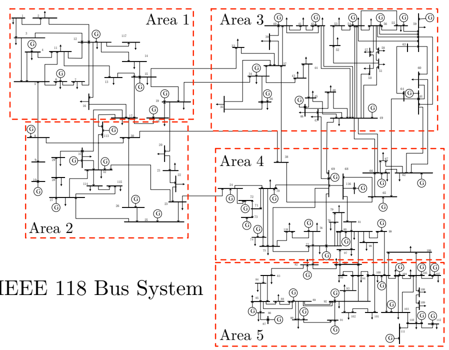




1. undetectable attacks exist
2. input/output (intruder/monitor) system has invariant zero
3. number of attacked signals > size of input/output linking

Result #2: Distributed Monitor Design

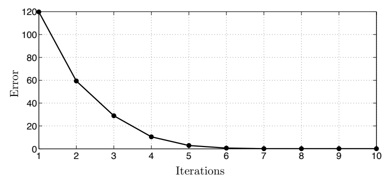
IEEE 118 bus (Midwest, 54-m 118-b)



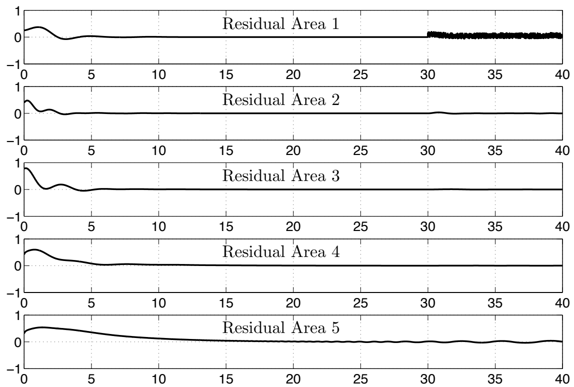
IEEE 118 Bus System

- Detection via residual filter design
- Centralized and distributed filters
- Distributed iterative filters via waveform relaxation

Waveform iteration error:

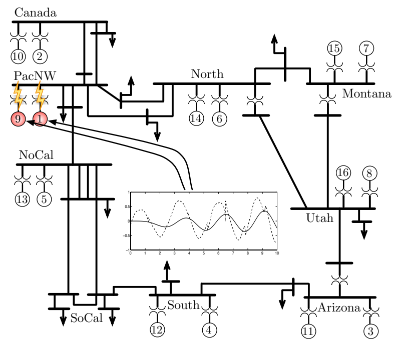


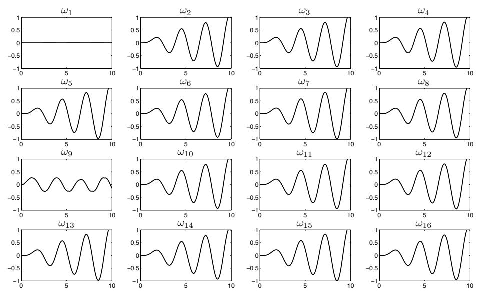
Residuals $r_i^{(k)}(t)$ for $k = 100$:



Result #3: Optimal Cooperative Attacks

Western US (WECC, 16-m 13-b)





- Optimal attack design via geometric control
- Two attackers suffice for network-wide instability
- Specific effect against selected machines
- Attack unidentifiable by single machine

De Marco et al, "Malicious control in a competitive power systems environment" CCA '96

Outline

Technical Assumptions

$$E\dot{x}(t) = Ax(t) + B_K u_K(t)$$
$$y(t) = Cx(t) + D_K u_K(t)$$

Technical assumptions guaranteeing existence, uniqueness, & smoothness:

(i) (E, A) is regular: $|sE - A|$ does not vanish for all $s \in \mathbb{C}$

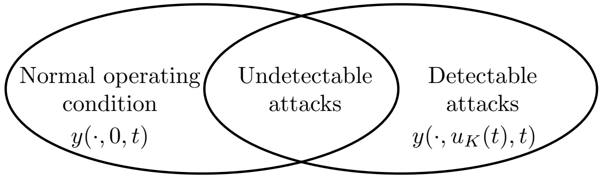
(ii) the initial condition $x(0)$ is consistent (can be relaxed)

(iii) the unknown input $u_K(t)$ is sufficiently smooth (can be relaxed)

- Attack set K = sparsity pattern of attack input

Undetectable Attack
Definition

An attack remains undetected if its effect on measurements is undistinguishable from the effect of some nominal operating conditions



Definition (Undetectable attack set)

The attack set K is *undetectable* if there exist initial conditions x_1, x_2, and an attack mode $u_K(t)$ such that, for all times t

$$y(x_1, u_K, t) = y(x_2, 0, t).$$

Undetectable Attack
Condition

By linearity, an undetectable attack is such that $y(x_1 - x_2, u_K, t) = 0$

- zero dynamics

Theorem

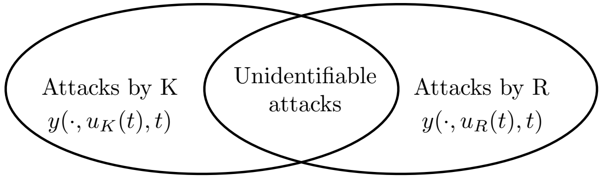
For the attack set K, there exists an undetectable attack if and only if

$$\begin{bmatrix} sE - A & -B_K \\ C & D_K \end{bmatrix} \begin{bmatrix} x \\ g \end{bmatrix} = 0$$

for some s, $x \neq 0$, and g.

Unidentifiable Attack
Definition

The attack set K remains unidentified if its effect on measurements is undistinguishable from an attack generated by a distinct attack set $R \neq K$



Definition (Unidentifiable attack set)

The attack set K is *unidentifiable* if there exists an admissible attack set $R \neq K$such that

$$y(x_K, u_K, t) = y(x_R, u_R, t).$$

Unidentifiable Attack
Condition

By linearity, the attack set K is unidentifiable if and only if there exists a distinct set $R \neq K$ such that $y(x_K - x_R, u_K - u_R, t) = 0$.

Theorem

For the attack set K, there exists an unidentifiable attack if and only if

$$\begin{bmatrix} sE - A & -B_K & -B_R \\ C & D_K & D_R \end{bmatrix} \begin{bmatrix} x \\ g_K \\ g_R \end{bmatrix} = 0$$

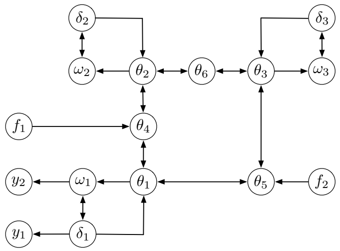
for some s, $x \neq 0$, g_K, and g_R.

So far we have shown:

- fundamental detection/identification limitations
- system-theoretic conditions for undetectable/unidentifiable attacks

From Algebraic to Graph-theoretical Conditions

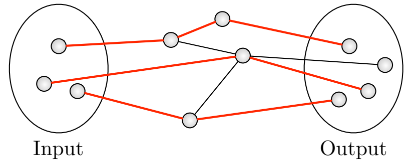
$$E\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$



- the vertex set is the union of the state, input, and output variables
- edges corresponds to nonzero entries in E, A, B, C, and D

Zero Dynamics and Connectivity

A linking between two sets of vertices is a set of mutually-disjoint directed paths between nodes in the sets



Theorem (Detectability, identifiability, linkings, and connectivity)

If the maximum size of an input-output linking is k:

- *there exists an undetectable attack set K_1, with $|K_1| \geq k$, and*
- *there exists an unidentifiable attack set K_2, with $|K_2| \geq \lceil \frac{k}{2} \rceil$.*

- statement becomes necessary with *generic* parameters
- statement applies to systems with parameters in polytopes

Summary and Future Directions

Cyber-Physical Security

1. fundamental limitations
2. distributed monitor design
3. control theory + distributed algorithms

Research Avenues

1. optimal network clustering for distributed procedures
2. analysis of costs and effects of attacks
3. optimal monitors with noise and faults
4. nonlinear and piecewise systems
5. integration with hypothesis testing and system optimization