Cyber-Physical Systems under Attack

Models, Fundamental limitations, and Monitor Design

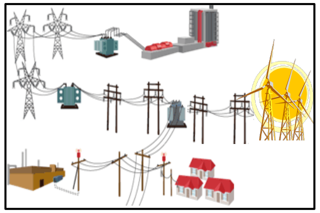
Fabio Pasqualetti

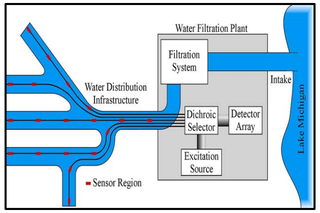
Florian Dörfler   Francesco Bullo

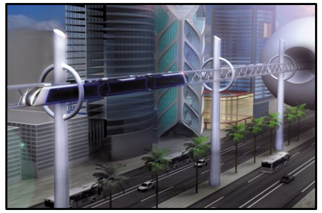
Center for Control, Dynamical systems and Computation
University of California, Santa Barbara

Workshop on Control Systems Security: Challenges and Directions
IEEE CDC, Orlando, FL, Dec 11, 2011

Important Examples of Cyber-Physical Systems







Many critical infrastructures are cyber-physical systems:

- power generation and distribution networks
- water networks and mass transportation systems
- econometric models (W. Leontief, *Input - output economics*, 1986)
- sensor networks
- energy-efficient buildings (heat transfer)

Security and Reliability of Cyber-Physical Systems

Cyber-physical security is a fundamental obstacle challenging the smart grid vision.

H. Khurana, "Cybersecurity: A key smart grid priority," *IEEE Smart Grid Newsletter*, Aug. 2011.

J. Meserve "Sources: Staged cyber attack reveals vulnerability in power grid" *http://cnn.com*, 2007.

A. R. Metke and R. L. Ekl "Security technology for smart grid networks," *IEEE Transactions on Smart Grid*, 2010.

J. P. Farwell and R. Rohozinski "Stuxnet and the Future of Cyber War" *Survival*, 2011.

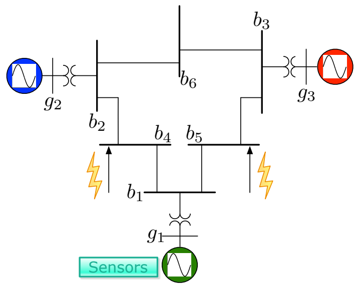
T. M. Chen and S. Abu-Nimeh "Lessons from Stuxnet" *Computer*, 2011.

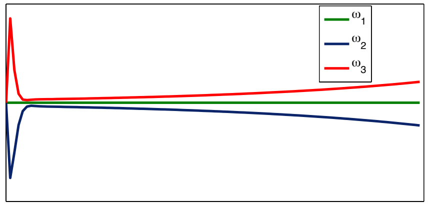
Water supply networks are among the nation's most critical infrastructures

J. Slay and M. Miller. "Lessons learned from the Maroochy water breach" *Critical Infrastructure Protection*, 2007.

D. G. Eliades and M. M. Polycarpou. "A Fault Diagnosis and Security Framework for Water Systems" *IEEE Transactions on Control Systems Technology*, 2010.

A Simple Example: WECC 3-machine 6-bus System





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

From Fault Detection and Cyber Security to Cyber-Physical Security

Cyber-physical security exploits system dynamics to assess correctness of measurements, and compatibility of measurement equation

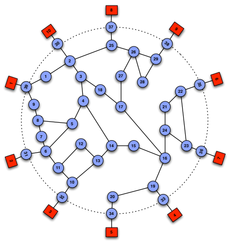
Cyber-physical security extends classical fault detection, and complements/augments cyber security

- classical fault detection considers only *generic* failures, while cyber-physical attacks are worst-case attacks
- cyber security does not exploit compatibility of measurement data with physics/dynamics
- cyber security methods are ineffective against attacks that affect the physics/dynamics

Models of Cyber-Physical Systems: Power Networks

Small-signal structure-preserving power network model:

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

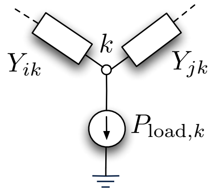


2. generators ■ modeled by swing equations:

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

3. buses ● with constant real power demand:

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

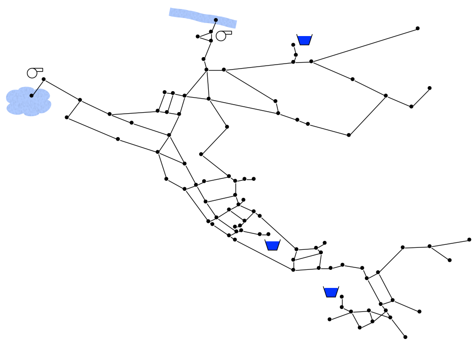


⇒ Linear differential-algebraic dynamics: $E\dot{x} = Ax$

Models of Cyber-physical Systems: 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: $E\dot{x} = Ax$

Models for Attackers and Security System

Byzantine Cyber-Physical Attackers

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

Security System

1. knows structure and parameters
2. measures output signal
3. security systems's objective is to detect and identify attack

1. characterize fundamental limitations on security system
2. design filters for detectable and identifiable attacks

Model of Cyber-Physical Systems under Attack

1. **Physics** obey linear differential-algebraic dynamics: $E\dot{x}(t) = Ax(t)$
2. **Measurements** are in continuous-time: $y(t) = Cx(t)$
3. **Cyber-physical attacks** are modeled as unknown input $u(t)$ with unknown input matrices B & D

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

This model includes **genuine faults** of system components, **physical attacks**, and **cyber attacks** caused by an omniscient malicious intruder.

Q: Is the attack $\big(B, D, u(t)\big)$ detectable/identifiable from the output $y(t)$?

Related Results on Cyber-Physical Security

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.

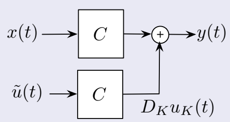
R. Smith, "A decoupled feedback structure for covertly appropriating network control systems," *IFAC World Congress*, Aug. 2011.

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.

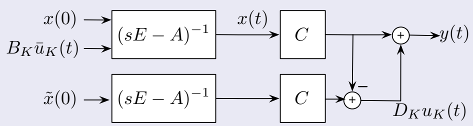
Our framework includes and generalizes most of these results

Prototypical Attacks

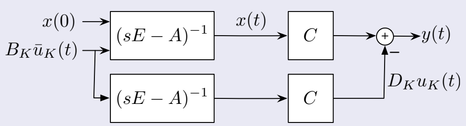
Static stealth attack:
corrupt measurements according to C



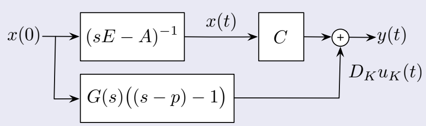
Replay attack:
effect system and reset output



Covert attack:
closed loop replay attack



Dynamic false data injection:
render unstable pole unobservable



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 $s \in \mathbb{C}$

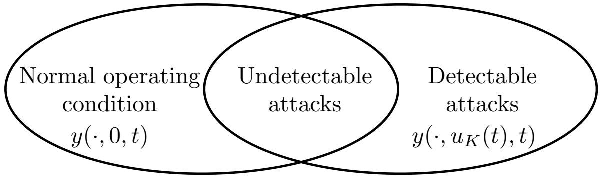
(ii) the initial condition $x(0)$ is consistent (can be relaxed)

(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

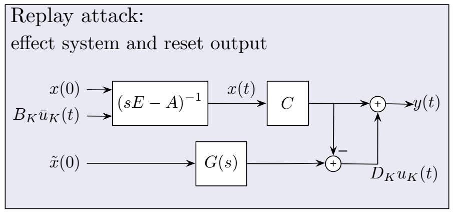
Theorem

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.

Undetectability of Replay Attacks



1. two attack channels: $\bar{u}_K$, u_K
2. $\mathrm{Im}(C) \subseteq \mathrm{Im}(D_K)$
3. $B_K \neq 0$

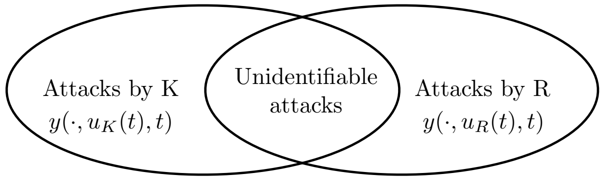
Undetectability follows from solvability of

$$\begin{bmatrix} sE - A & -B_K & 0 \\ C & 0 & D_K \end{bmatrix} \begin{bmatrix} x \\ g_1 \\ g_2 \end{bmatrix} = 0$$

- $x = (sE - A)^{-1} B_K g_1$, $g_2 = D_K^{\dagger} C (sE - A)^{-1} B_K g_1$
- replay attacks can be detected though *active detectors*
- replay attacks are not worst-case attacks

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

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

- an undetectable attack set is also unidentifiable

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

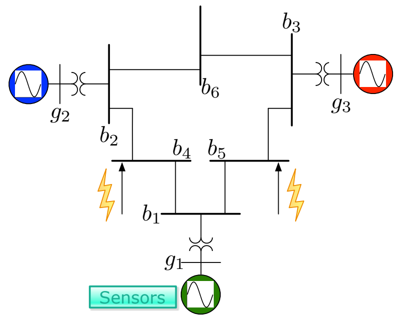
$$\begin{bmatrix} sE - A & -B_K & -B_R \\ C & D_K & D_R \end{bmatrix} \begin{bmatrix} x \\ g_K \\ g_R \end{bmatrix} = 0$$

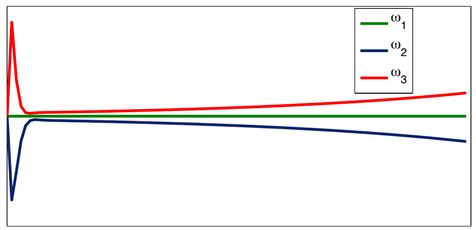
for some s, $x \neq 0$, g_K, and g_R.

So far we have shown:

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

WECC 3-machine 6-bus System



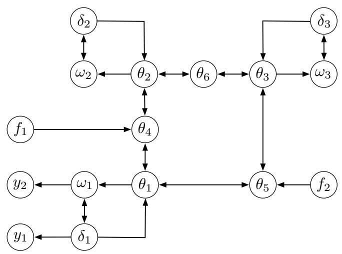


1. **Physical dynamics:** classical generator model & DC load flow
2. **Measurements:** angle and frequency of generator g_1
3. **Attack:** modified real power injections at buses b_4 & b_5

The attack through b_4 and b_5 excites only zero dynamics for the measurements at the first generator

From Algebraic to Graph-theoretical Conditions

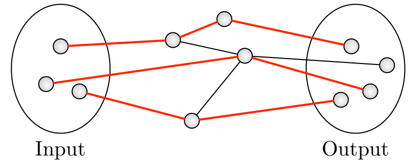
$$E\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$



- 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



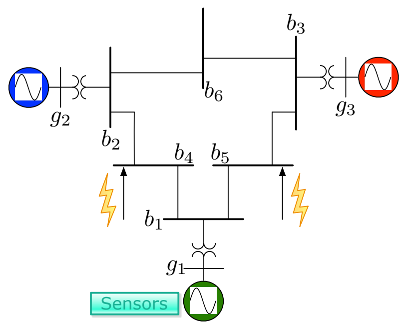
Theorem (Detectability, identifiability, linkings, and connectivity)

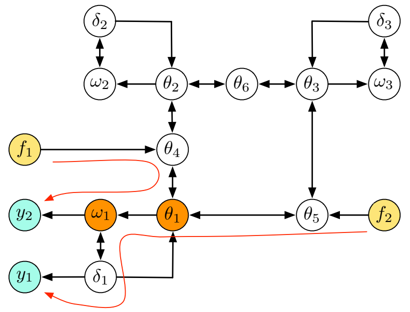
If the maximum size of an input-output linking is k:

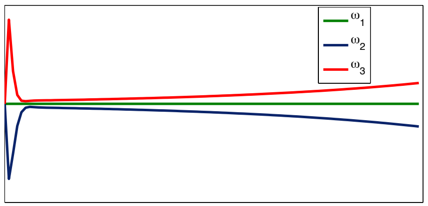
- *there exists an undetectable attack set K_1, with $|K_1| \geq k$, and*
- *there exists an unidentifiable attack set K_2, with $|K_2| \geq \lceil \frac{k}{2} \rceil$.*

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

WECC 3-machine 6-bus System Revisited







1. #attacks > max size linking
2. $\exists$ undetectable attacks
3. attack destabilizes g_2, g_3

Centralized Detection Monitor Design

System under attack $(B, D, u(t))$:

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

Proposed centralized detection filter:

$$E\dot{w}(t) = (A + GC)w(t) - Gy(t)$$
$$r(t) = Cw(t) - y(t)$$

Theorem (Centralized Attack Detection Filter)

Assume $w(0) = x(0)$, $(E, A + GC)$ is Hurwitz, and attack is detectable. Then $r(t) = 0$ if and only if $u(t) = 0$.

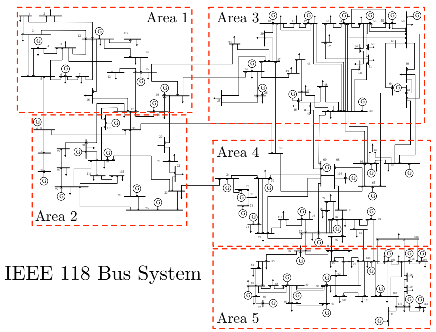
- ☺ the design is independent of B, D, and $u(t)$
- ☺ if $w(0) \neq x(0)$, then asymptotic convergence
- ☹ a direct centralized implementation may not be feasible
  due to high-dimensionality of a power network, communication complexity, ...

Decentralized Monitor Design

Partition the physical system with geographically deployed control centers:

$$E = \begin{bmatrix} E_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & E_N \end{bmatrix}, \quad C = \begin{bmatrix} C_1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & C_N \end{bmatrix}$$

$$A = \begin{bmatrix} A_1 & \cdots & A_{1N} \\ \vdots & \vdots & \vdots \\ A_{N1} & \cdots & A_N \end{bmatrix} = A_D + A_C$$



(i) control center i knows E_i, A_i, and C_i, and neighboring A_{ij}

(ii) control center i can communicate with control center $j \Leftrightarrow A_{ji} \neq 0$

(iii) E&C are blockdiagonal, (E_i, A_i) is regular & (E_i, A_i, C_i) is observable

Decentralized Monitor Design: Continuous Communication

System under attack:

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

where $A = A_D + A_C$

Decentralized detection filter:

$$E\dot{w}(t) = (A_D + GC)w(t) + A_C w(t) - Gy(t)$$
$$r(t) = Cw(t) - y(t)$$

where $G = \mathrm{blkdiag}(G_1, \ldots, G_N)$

Theorem (Decentralized Attack Detection Filter)

Assume that $w(0) = x(0)$, $(E, A_D + GC)$ is Hurwitz, and

$$\rho\left((j\omega E - A_D - GC)^{-1} A_C\right) < 1 \quad \textit{for all } \omega \in \mathbb{R}.$$

If the attack is detectable, then $r(t) = 0$ if and only if $u(t) = 0$.

- ☺ the design is decentralized but achieves centralized performance
- ☹ the design requires continuous communication among control centers

Digression: Gauss-Jacobi Waveform Relaxation

- **Standard Gauss-Jacobi relaxation** to solve a linear system $Ax = u$:

$$x_i^{(k)} = \frac{1}{a_{ii}}\Big(u_i - \sum\nolimits_{j\neq i} a_{ij} x_j^{(k-1)}\Big) \quad \Leftrightarrow \quad x^{(k)} = -A_D^{-1} A_C x^{(k-1)} + A_D^{-1} u$$

Convergence: $\lim_{k\to\infty} x^{(k)} \to x = A^{-1}u \quad \Leftrightarrow \quad \boxed{\rho(A_D^{-1}A_C) < 1}$

- **Gauss-Jacobi waveform relaxation** to solve $E\dot{x}(t) = Ax(t) + Bu(t)$:

$$E\dot{x}^{(k)}(t) = A_D x^{(k)}(t) + A_C x^{(k-1)}(t) + Bu(t), \quad t \in [0, T]$$

Convergence for (E, A) Hurwitz & $u(t)$ integrable in $t \in [0, T]$:

$$\lim_{k\to\infty} x^{(k)}(t) \to x(t) \quad \Leftarrow \quad \boxed{\rho\left((j\omega E - A_D)^{-1} A_C\right) < 1 \quad \forall\, \omega \in \mathbb{R}}$$

Distributed Monitor Design: Discrete Communication

Distributed attack detection filter:

$$\begin{aligned} E\dot{w}^{(k)}(t) &= (A_D + GC) w^{(k)}(t) + A_C w^{(k-1)}(t) - Gy(t) \\ r^{(k)}(t) &= Cw^{(k)}(t) - y(t) \end{aligned}$$

where $G = \text{blkdiag}(G_1, \ldots, G_N)$, $t \in [0, T]$, and $k \in \mathbb{N}$

Theorem (Distributed Attack Detection Filter)

Assume that $w^{(k)}(0) = x(0)$ for all $k \in \mathbb{N}$, $y(t)$ is integrable for $t \in [0, T]$, $(E, A_D + GC)$ is Hurwitz, and

$$\rho\left((j\omega E - A_D - GC)^{-1} A_C\right) < 1 \quad \text{for all } \omega \in \mathbb{R}.$$

If the attack is detectable, then $\lim_{k\to\infty} r^{(k)}(t) = 0$ if and only if $u(t) = 0$ for all $t \in [0, T]$.

Implementation of Distributed Attack Detection Filter

Distributed iterative procedure to compute the residual $r(t)$, $t \in [0, T]$:

1. set $k := k + 1$, and compute $w_i^{(k)}(t)$, $t \in [0, T]$, by integrating

$$E_i \dot{w}_i^{(k)}(t) = (A_i + G_i C_i) w_i^{(k)}(t) + \sum\nolimits_{j\neq i} A_{ij} w_j^{(k-1)}(t) - G_i y_i(t)$$

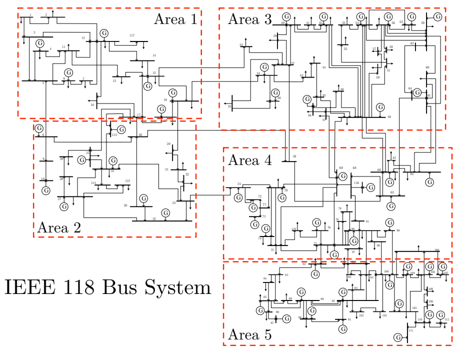
2. transmit $w_i^{(k)}(t)$ to control center j if $A_{ij} \neq 0$
3. update $w_j^{(k)}(t)$ with the signal received from control center j

$\Rightarrow$ For k sufficiently large, $r_i^{(k)}(t) = C_i w_i^{(k)}(t) - y_i(t) \approx 0 \;\Leftrightarrow\;$ no attack

$\Rightarrow$ Receding horizon implementation: move integration window $[0, T]$

$\Rightarrow$ Distributed verification of convergence cond.: $\rho(\cdot) < 1 \;\Leftarrow\; \|\cdot\|_\infty < 1$.

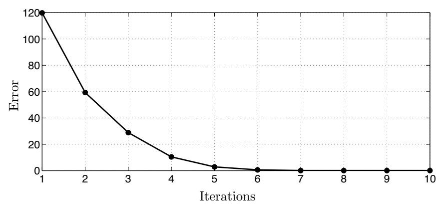
An Illustrative Example: IEEE 118 Bus System



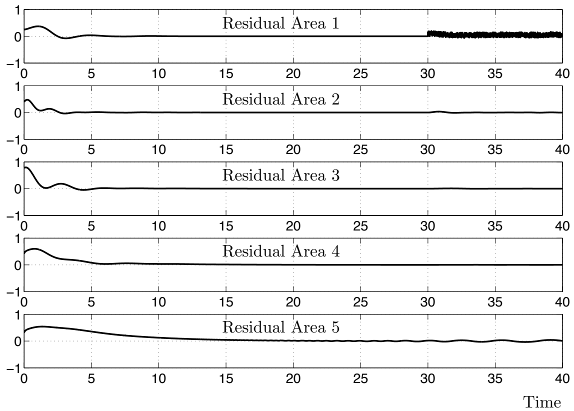
IEEE 118 Bus System

- **Physics:** classical generator model and DC load flow model
- **Measurements:** generator angles
- **Attack** of all measurements in Area 1

Convergence of waveform relaxation:



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



Centralized Identification Monitor Design

System under attack $(B_K, D_K, u_K(t))$:

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

Centralized identification filter:

$$E\dot{w}(t) = \bar{A}w(t) - \bar{G}y(t)$$
$$r_K(t) = MCw(t) - Hy(t)$$

- only $u_K(t)$ is active, i.e., $u_R(t) = 0$ at all times

Theorem

Assume $w(0) = x(0)$, and attack set is identifiable.
Then $r_K(t) = 0$ if and only if K is the attack set.

- ☺ if $w(0) \neq x(0)$, then asymptotic convergence
- ☹ a direct centralized implementation may not be feasible
- ☹ design depends on $(B_K, D_K) \Rightarrow$ combinatorial complexity (NP-hard)

Design Method

Controlled, Conditioned, and Deflating Subspaces

Let $\mathcal{S}_K^*$ be the smallest subspace of the state space such that

- $\exists\ G$ such that $(A + GC)\mathcal{S}_K^* \subseteq \mathcal{S}_K^*$ and $\mathcal{R}(B_K + GD_K) \subseteq \mathcal{S}_K^*$

Design steps:

- compute smallest conditioned invariant subspace $\mathcal{S}_K$
- make the subspace $\mathcal{S}_K$ invariant by output injection
- build a residual generator for the quotient space $\mathcal{X} \setminus \mathcal{S}_K^*$
- the residual is not affected by $u_K(t)$

Conclusion

We have presented:

1. a modeling framework for cyber-physical systems under attack
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

Ongoing and future work:

1. optimal **network partitioning** for distributed procedures
2. effect of **noise**, modeling uncertainties & communication constraints
3. quantitative analysis of **cost** and **effect** of attacks
4. applications to distributed-parameters cyber-physical systems

References

F. Pasqualetti, A. Bicchi, and F. Bullo. Distributed intrusion detection for secure consensus computations. In *IEEE Conf. on Decision and Control*, pages 5594–5599, New Orleans, LA, USA, Dec. 2007.

F. Pasqualetti, A. Bicchi, and F. Bullo. On the security of linear consensus networks. In *IEEE Conf. on Decision and Control and Chinese Control Conference*, pages 4894–4901, Shanghai, China, Dec. 2009.

F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 2011, DOI: 10.1109/TAC.2011.2158130.

F. Pasqualetti, R. Carli, A. Bicchi, and F. Bullo. Identifying cyber attacks under local model information. In *IEEE Conf. on Decision and Control*, Atlanta, GA, USA, December 2010.

F. Pasqualetti, R. Carli, A. Bicchi, and F. Bullo. Distributed estimation and detection under local information. In *IFAC Workshop on Distributed Estimation and Control in Networked Systems*, Annecy, France, September 2010.

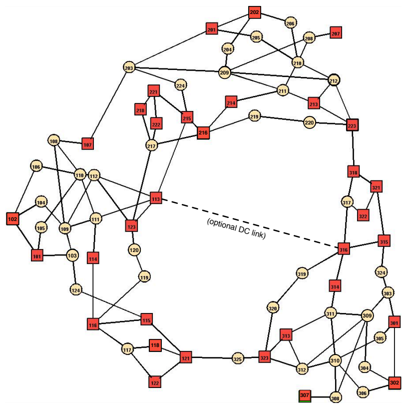
F. Pasqualetti, A. Bicchi, and F. Bullo. A graph-theoretical characterization of power network vulnerabilities. In *American Control Conference*, San Francisco, CA, USA, June 2011.

F. Pasqualetti, R. Carli, and F. Bullo. Distributed estimation and false data detection with application to power networks. *Automatica*, March 2011, To appear.

F. Pasqualetti, F. Dörfler, and F. Bullo. Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design. In *IEEE Conf. on Decision and Control*, Orlando, FL, USA, December 2011. To appear

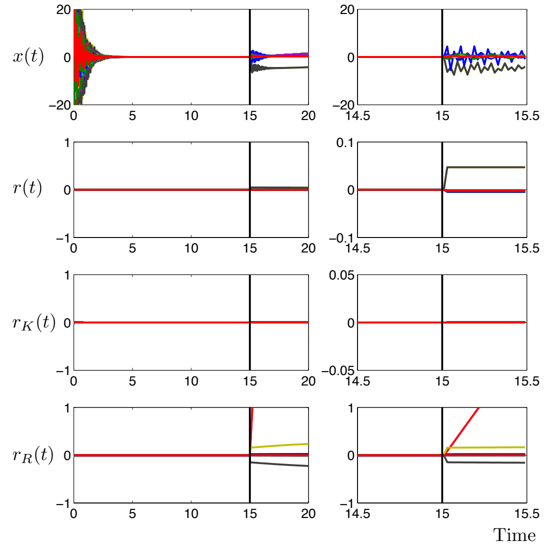
F. Dörfler, F. Pasqualetti, and F. Bullo. "Distributed detection of cyber-physical attacks in power networks: A waveform relaxation approach," in *Allerton Conf. on Communications, Control and Computing*, Sep. 2011.

A Case Study: RTS-96 Bus System



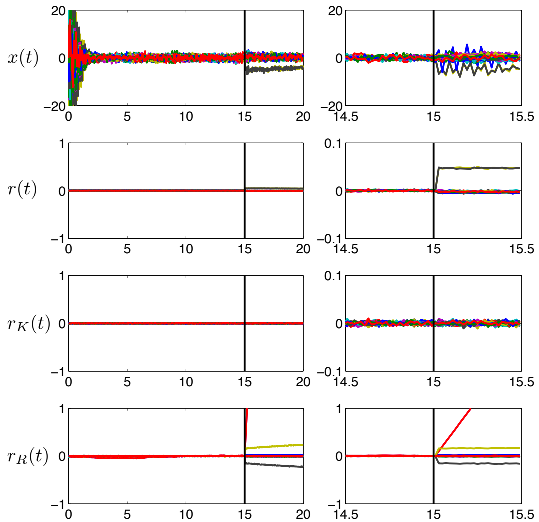
1. **Physical dynamics:** classical generator model & DC load flow
2. **Measurements:** angle and frequency of all generators
3. **Attack:** modify mechanical power injections at generators g_{101} & g_{102}
4. **Monitors:** our centralized detection and identification filters

RTS-96 Bus System: Linear Dynamics without Noise



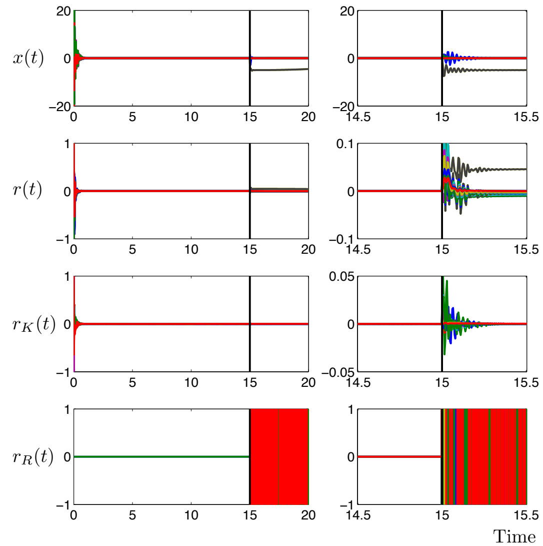
- $x(t)$: generators trajectories
- $r(t)$: detection residual
- $r_K(t)$: identification residual for K
- $r_R(t)$: identification residual for R
- filters are designed via conditioned invariance technique

RTS-96 Bus System: Linear Dynamics with Noise



- $x(t)$: generators trajectories
- $r(t)$: detection residual
- $r_K(t)$: identification residual for K
- $r_R(t)$: identification residual for R
- filters are designed via conditioned invariance and Kalman gain

RTS-96 Bus System: Nonlinear Dynamics



- $x(t)$: generators trajectories
- $r(t)$: detection residual
- $r_K(t)$: identification residual for K
- $r_R(t)$: identification residual for R
- filters are designed via conditioned invariance and Kalman gain