Short Course:
Topics on Cyber-Physical Control Systems

Karl H. Johansson
ACCESS Linnaeus Center & School of Electrical Engineering
KTH Royal Institute of Technology, Sweden

Slides and papers available at http://people.kth.se/~kallej

Department of Electronic & Computer Engineering
Hong Kong University of Science and Technology, July 2015

Course Outline

Jul 20: What is a cyber-physical system?

Jul 20: Event-based control of networked systems

Jul 22: Cyber-secure networked control systems

Aug 5: IAS Lecture on "Cyber-physical control for sustainable freight transportation"

Cyber-secure networked control systems

Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

Acknowledgements

Presentation based on joint papers with

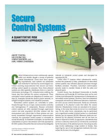
Henrik Sandberg (KTH)

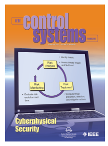
André Teixeira (KTH, soon TU Delft)

Kin C. Sou (Chalmers)

Iman Shames (U Melbourne)

Julien M. Hendrickx, Raphaël M. Jungers (UC Louvain)





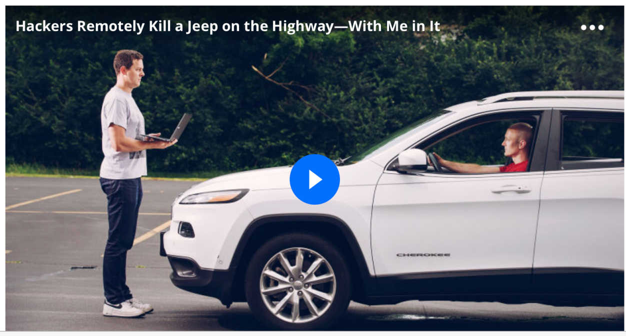
Funding sources:

WIRED | Hackers Remotely Kill a Jeep on the Highway—With Me in It | SUBSCRIBE

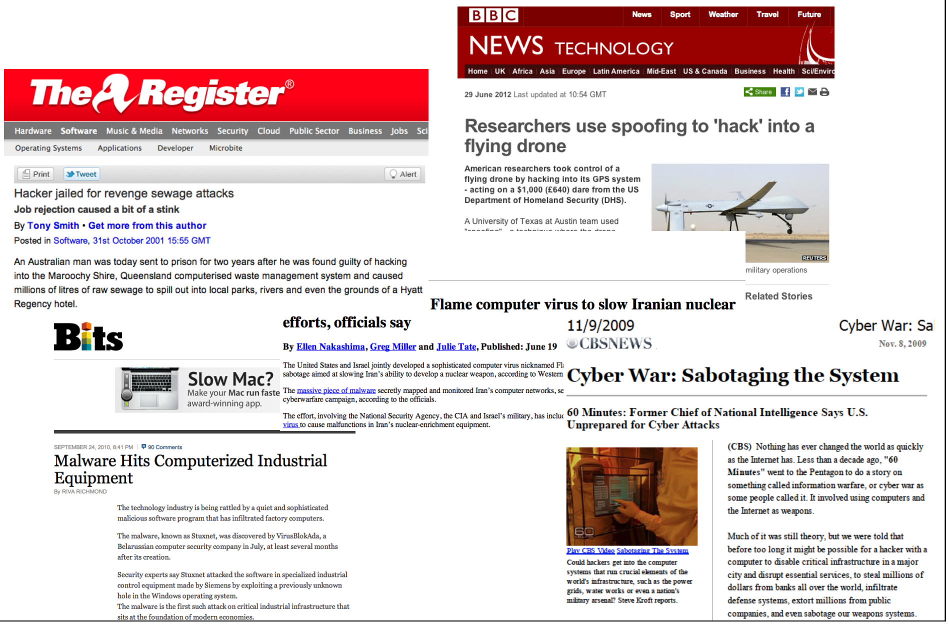
BUSINESS | DESIGN | ENTERTAINMENT | GEAR | SCIENCE | SECURITY

ANDY GREENBERG SECURITY 07.21.15 6:00 AM

HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT



Some Other Cyber-Attacks on Control Systems


The Register
Hardware Software Music & Media Networks Security Cloud Public Sector Business Jobs
Operating Systems Applications Developer Microbite

Hacker jailed for revenge sewage attacks
Job rejection caused a bit of a stink
By Tony Smith • Get more from this author
Posted in Software, 31st October 2001 15:55 GMT

An Australian man was today sent to prison for two years after he was found guilty of hacking into the Maroochy Shire, Queensland computerised waste management system and caused millions of litres of raw sewage to spill out into local parks, rivers and even the grounds of a Hyatt Regency hotel.

BBC NEWS TECHNOLOGY
29 June 2012 Last updated at 10:54 GMT

Researchers use spoofing to 'hack' into a flying drone

American researchers took control of a flying drone by hacking into its GPS system - acting on a $1,000 (£640) dare from the US Department of Homeland Security (DHS).

A University of Texas at Austin team used

Flame computer virus to slow Iranian nuclear efforts, officials say

By Ellen Nakashima, Greg Miller and Julie Tate, Published: June 19

Bits

Malware Hits Computerized Industrial Equipment

The technology industry is being rattled by a quiet and sophisticated malicious software program that has infiltrated factory computers.

The malware, known as Stuxnet, was discovered by VirusBlokAda, a Belarussian computer security company in July, at least several months after its creation.

Security experts say Stuxnet attacked the software in specialized industrial control equipment made by Siemens by exploiting a previously unknown hole in the Windows operating system.

The malware is the first such attack on critical industrial infrastructure that sits at the foundation of modern economies.

11/9/2009
CBSNEWS

Cyber War: Sabotaging the System

60 Minutes: Former Chief of National Intelligence Says U.S. Unprepared for Cyber Attacks

(CBS) Nothing has ever changed the world as quickly as the Internet has. Less than a decade ago, "60 Minutes" went to the Pentagon to do a story on something called information warfare, or cyber war as some people called it. It involved using computers and the Internet as weapons.

Much of it was still theory, but we were told that before too long it might be possible for a hacker with a computer to disable critical infrastructure in a major city and disrupt essential services, to steal millions of dollars from banks all over the world, infiltrate defense systems, extort millions from public companies, and even sabotage our weapons systems.

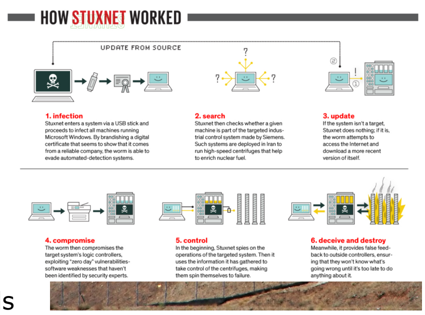
The Stuxnet Worm 2010

Targets: MS Windows, programmable logic controllers, industrial control system, connected to variable-frequency drives

Exploited **4 zero-day flaws** (security holes not known to vendor)

Speculated goal:
Harm centrifuges at uranium enrichment facility in Iran

Attack mode:
1. Delivery with USB stick (**no internet connection**)
2. Replay measurements to control center and execute harmful controls



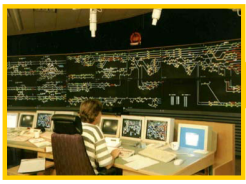
["The Real Story of Stuxnet", IEEE Spectrum, 2013]

Motivation

- Northeast blackout Aug 14, 2003: 55 million people affected
- Software bug in energy management system **stalled alarms in state estimator for over an hour**
- Cyber-attacks against the power network control systems with similar consequences pose a substantial threat

From security requirements to societal cost

Attack



SCADA system

Power network

Societal cost

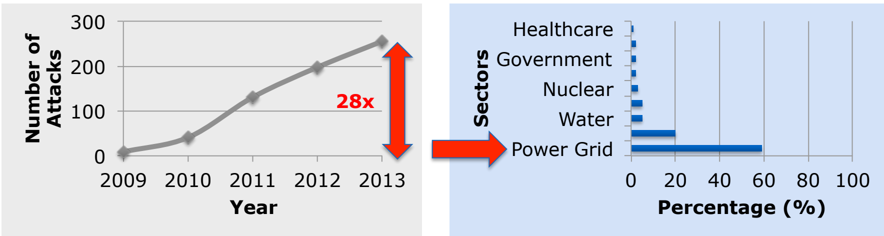
Security issues

Power system: susceptible to operational errors and external attacks

Smart grid technology makes the system even more vulnerable

SEVENTH FRAMEWORK PROGRAMME VIKING

Cyber Incidents in US Critical Infrastructures



ICS-CERT = Industrial Control Systems Cyber Emergency Response Team, https://ics-cert.us-cert.gov, US Department of Homeland Security

[ICS-CERT, 2013; Zonouz, 2014]

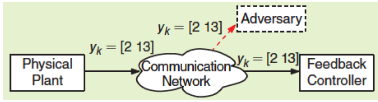
Information Security

Confidentiality: information is not disclosed to unauthorized individuals

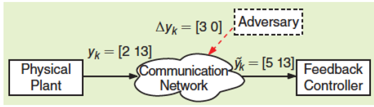
Integrity: information cannot be modified in an unauthorized manner

Availability: information must be available when it is needed

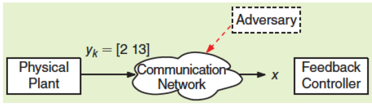
Control Systems Security



Confidentiality: information is not disclosed to unauthorized individuals



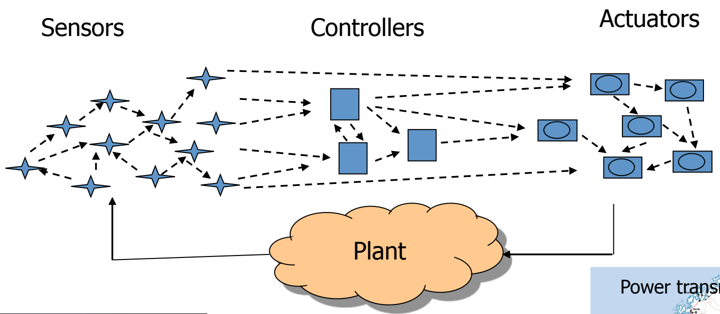
Integrity: information cannot be modified in an unauthorized manner

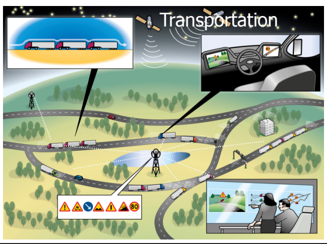


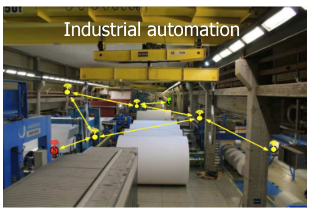
Availability: information must be available when it is needed

Integrity and availability are often the most critical security attributes for control systems

Networked Control Systems

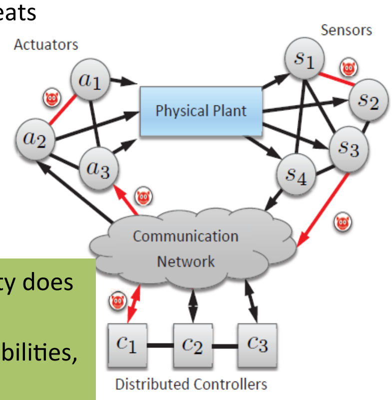






Cyber-Secure Networked Control Systems

- Networked control systems are to a growing extent based on **open communication and software technology**
- Leads to **increased vulnerability** to cyber-threats with many potential points of attacks



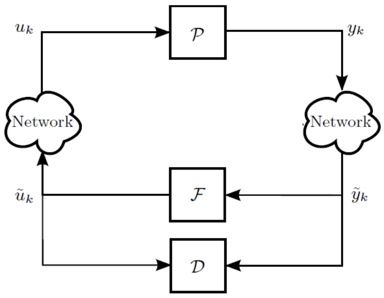
- How to model attacks?
- How to measure vulnerability?
- How to compute consequences?
- How to design protection mechanisms?

- Traditional computer and information security does not provide answers to these questions
- **Cyber-physical coupling** creates new vulnerabilities, but also new means for protection
- Infrastructure attacks can have dramatic impact

Outline

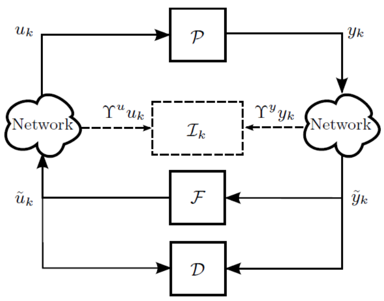
- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

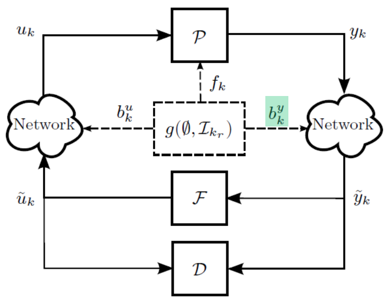
Networked Control System



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

Networked Control System under Attack



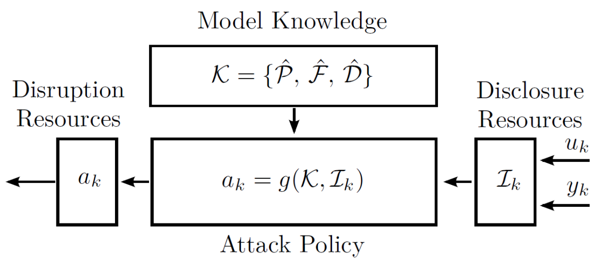


- Physical plant $\mathcal{P}$
- Feedback controller $\mathcal{F}$
- Anomaly detector $\mathcal{D}$

- Disclosure attack
- Physical attack f_k
- Deception attack

$$\tilde{u}_k = u_k + \Gamma^u b_k^u$$
$$\tilde{y}_k = y_k + \Gamma^y b_k^y$$

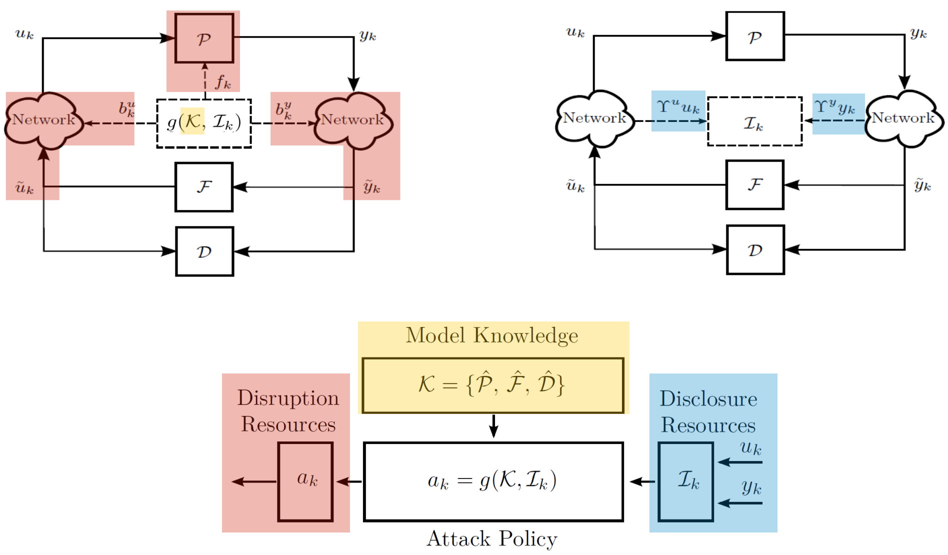
Adversary Model



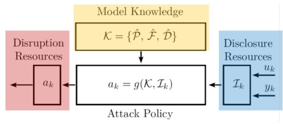
- Adversary constrained by limited resources
- Attack policy depends on adversary goals and constraints

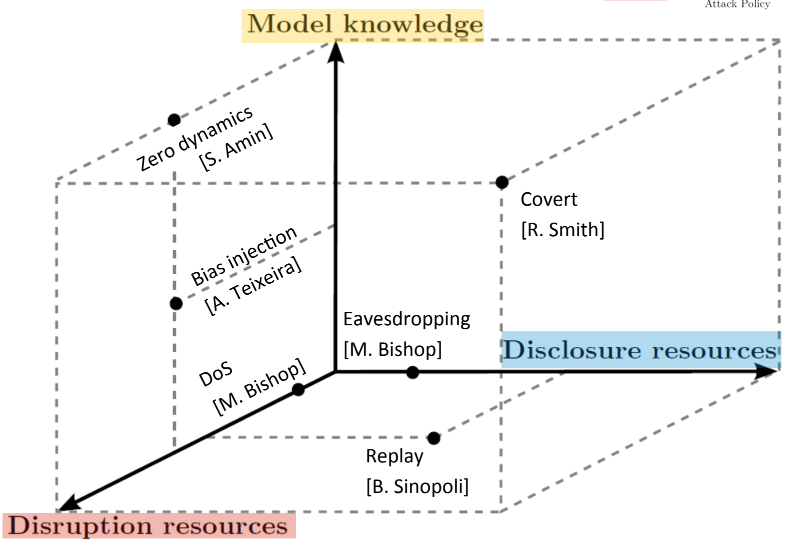
[Teixeira *et al.*, HiCoNS, 2012]

Networked Control System with Adversary Model



Attack Space



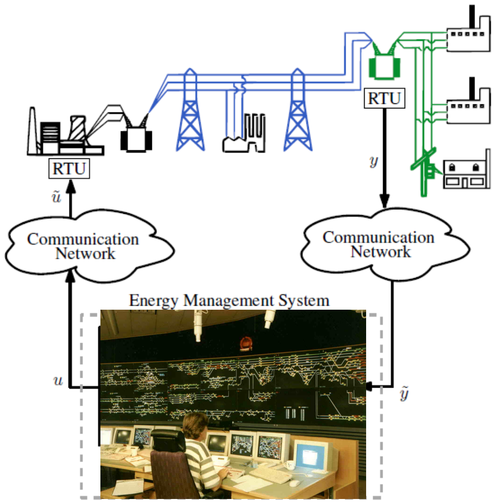


[Teixeira *et al.*, HiCoNS, 2012]

Outline

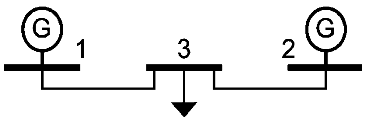
- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

Control of Transmission Power Network



(Static) Power Network Model

- Local states at bus i:
  - θ_i – phase angle
  - V_i – voltage magnitude



- Active and reactive power injections:

$$P_i = V_i \sum_{j \in N_i} V_j \left(G_{ij} \cos\theta_{ij} + B_{ij} \sin\theta_{ij}\right)$$
$$Q_i = V_i \sum_{j \in N_i} V_j \left(G_{ij} \sin\theta_{ij} - B_{ij} \cos\theta_{ij}\right)$$

- Active and reactive power flows:

$$P_{ij} = V_i^2 (g_{si} + g_{ij}) - V_i V_j \left(g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}\right)$$
$$Q_{ij} = -V_i^2 (b_{si} + b_{ij}) - V_i V_j \left(g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}\right)$$

where

$$\theta_{ij} = \theta_i - \theta_j$$

- Measurement model:

$$z = h(x) + \epsilon$$

  - $x \in \mathbb{R}^n$: network states
  - $z \in \mathbb{R}^m$: power flow measurements
  - ϵ : measurement noise

Static model because the power grid time constant ~10 ms is beyond existing measurement technology. Typical sampling time ~1 s.

Steady-State Power Flow Model

States θ
= bus voltage phase angles

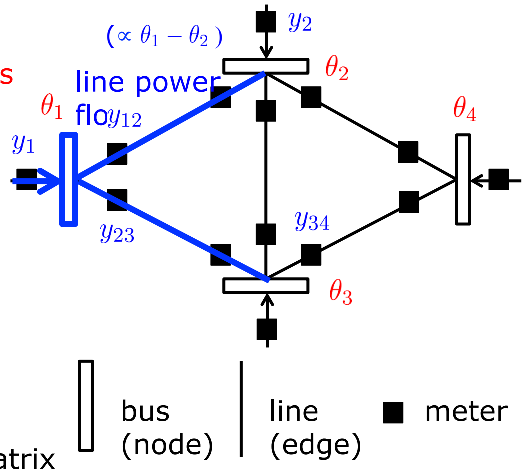
(flow conservation)
bus injection

Measurements y
= line power flow & bus injection

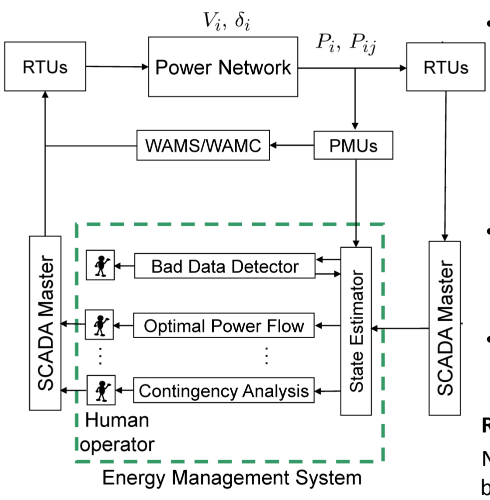
DC power flow model:

$$y = H\theta$$

measurement matrix



Energy Management System for Power Networks



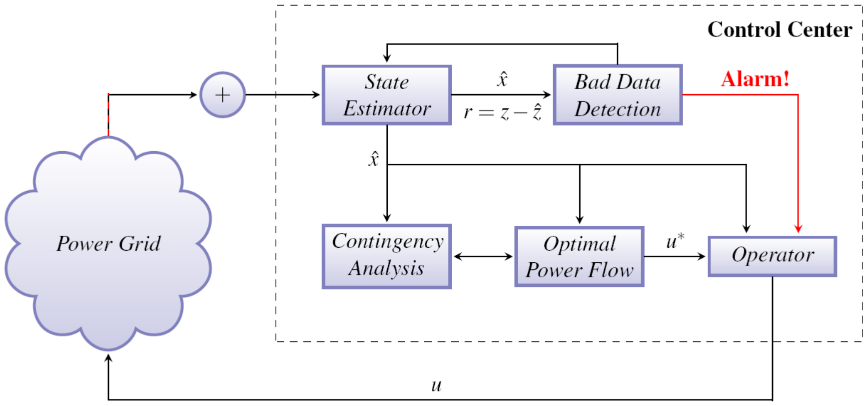
SCADA = Supervisory Control and Data Acquisition
WAMC = Wide Area Monitoring and Control System
RTUs = Remote Terminal Units (Sensors/Actuators)

- SCADA-EMS provides power network state information to
  - Identify faulty equipment
  - Optimize power flows
  - Analyze reliability (contingency)
  - Etc
- Large system with slow sampling
  - 100-1 000's of RTUs sampled in sec's
  - 10K-40K measurements
- Decisions taken by human operators

Remark
New WAMCs based on high-rate PMUs are better protected but constitute only a small portion of the overall network

PMUs = Phasor Measurement Units (Sensors)

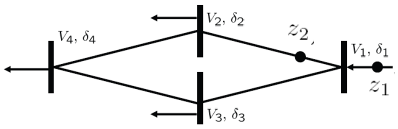
Energy Management System



- The **state estimator** has a crucial role in the EMS
- If the **bad data detector** identifies a faulty sensor, the corresponding measurement is removed from the state estimator
- Bad data detection is typically done under the assumption of **uncorrelated faults**, which does not hold for intelligent attacks

(Static) State Estimator

- Steady-state models:



$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} \frac{V_1 V_2}{X_{12}} \sin(\delta_1 - \delta_2) + \frac{V_1 V_3}{X_{13}} \sin(\delta_1 - \delta_3) \\ \frac{V_1 V_2}{X_{12}} \sin(\delta_1 - \delta_2) \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = h(x) + e \in \mathbb{R}^m$$

- WLS estimates of bus phase angles δ_i (in vector $\hat{x}$):

$$\hat{x}^{k+1} = \hat{x}^k + (H_k^T R^{-1} H_k)^{-1} H_k^T R^{-1} (z - h(\hat{x}^k))$$

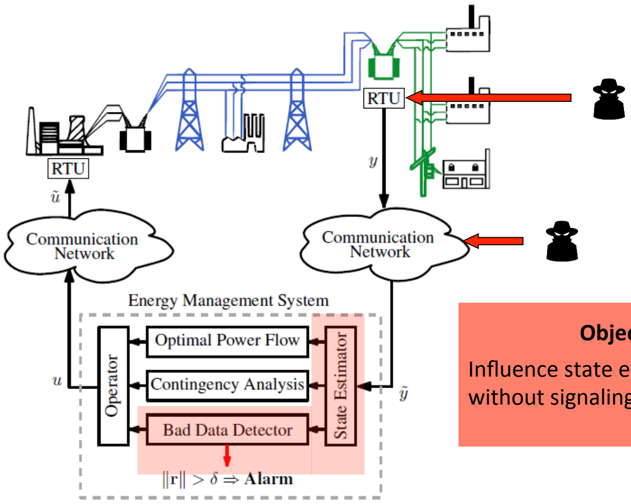
$$H_k := \frac{\partial h}{\partial x}(\hat{x}_k) \qquad R := \mathbf{E} e e^T$$

- Linear DC approximation (≈ ML estimate):

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} z \qquad H := \left. \frac{\partial h(x)}{\partial x} \right|_{x=0}$$

E.g., [Schweppe and Wildes, 1970; Abur and Exposito, 2004]

Attack Scenario

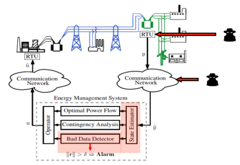


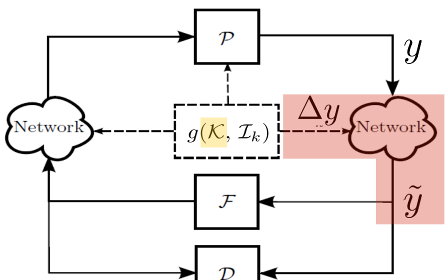
Objective

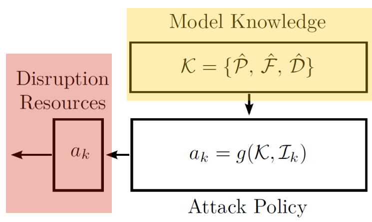
Influence state estimator without signaling bad data alarm

[Giani *et al.*, IEEE ISRCS, 2009; Mohajerin Esfahani *et al.*, CDC, 2010]

Adversary Model







- **Attack policy:** Induce bias in power measurements without alarms
- **Model knowledge:** Steady-state model of power system
- **Disruption resources:** Small number of measurement channels

How secure is the system against such an attack?

Structure of Measurement Matrix *H*

$$y = H\theta \quad \text{with} \quad H = \begin{bmatrix} DA^T \\ -DA^T \\ ADA^T \end{bmatrix} \begin{matrix} \text{(flow measurements)} \\ \text{(flow measurements)} \\ \text{(injection measurements)} \end{matrix}$$

- A - directed incidence matrix of power network
- D - diagonal matrix of reciprocals of transmission line reactance

Typically many more measurements than states

Data Influence on State Estimates

State estimator (LS)

$y = H\theta$

$\Rightarrow \hat{\theta} = (H^T H)^{-1} H^T y$

wrong

→ **wrong** Contingency analysis

→ **wrong** Optimal power flow

→ ⋮

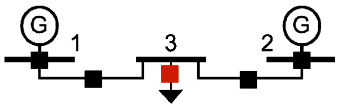
What if the measurements were **wrong**?

$\tilde{y} = y + \Delta y$ → random measurement noise

→ intentional data attack → $\tilde{\theta} = \hat{\theta} + \Delta\theta$

32

Example: Stealthy Attacks



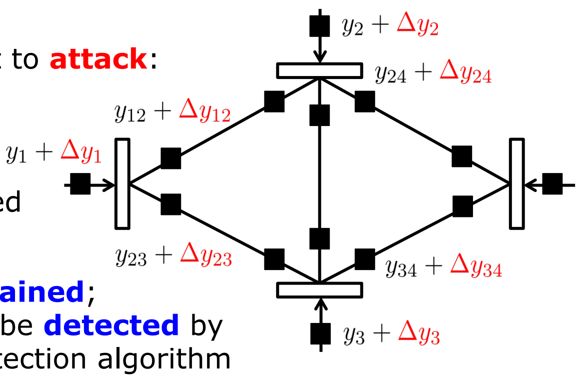
- P_3 is the target measurement
- A few possible attacks:
  - ~~$\{P_3\}, \{P_3, \star\}$~~ not stealthy
  - $\{P_1, P_{13}, P_3\}$ minimum effort
  - $\{P_2, P_{23}, P_3\}$
  - $\{P_1, P_{13}, P_3, P_{23}, P_2\}$

Stealthy Additive Deception Attack

Measurements subject to **attack**:

$\tilde{y} = y + \Delta y$

Is there a state explaining the received measurements?



Attack is **constrained**; otherwise it will be **detected** by the bad data detection algorithm

Stealth attack: $\Delta y = H \Delta \theta$

[Liu *et al.*, ACM CCCS, 2009; Sandberg *et al.*, CPSWEEK, 2010]

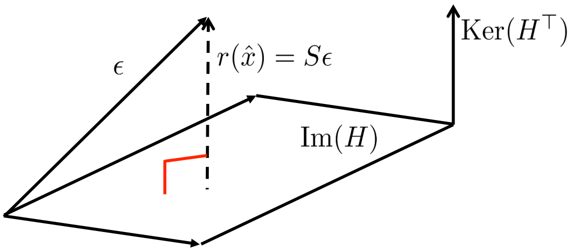
Geometric Interpretation of Bad Data Detection

$$H = \frac{\partial h(x)}{\partial x}\Big|_{x=\hat{x}}$$

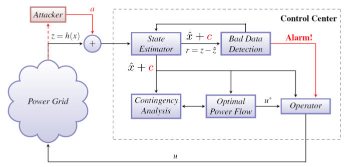
- Today's BDD is based on measurement residual $r(\hat{x}) = z - h(\hat{x})$

$$\|Wr(\hat{x})\|_p \underset{H_1}{\overset{H_0}{\lessgtr}} \tau$$

- For the Gauss-Newton method: $r(\hat{x}) \approx (I - H(H^\top H)^{-1} H^\top)\epsilon = S\epsilon$
- Note that $S = \mathbf{P}_{\mathrm{Ker}(H^\top)}$ is the orthogonal projection onto $\mathrm{Ker}(H^\top)$
- Can be exploited by an attacker



Attack Geometry



- Bad-data detection trigger alarm when residual r is large

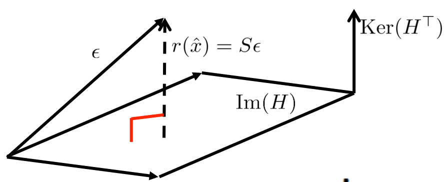
$$r := z - \hat{z} = z - H\hat{x} = z - H(H^T R^{-1} H)^{-1} H^T R^{-1} z$$

- Characterization of undetectable malicious data a

$$z_a := z + a$$

$$a = Hc \in \mathrm{Im}(H)$$

$$r = z - \hat{z} = z_a - \hat{z}_a$$



- The attacker has a lot of freedom in the choice of a!
- Attacker likely to seek sparse solutions a , i.e., manipulate only few measurements

[Liu *et al.*, 2009]

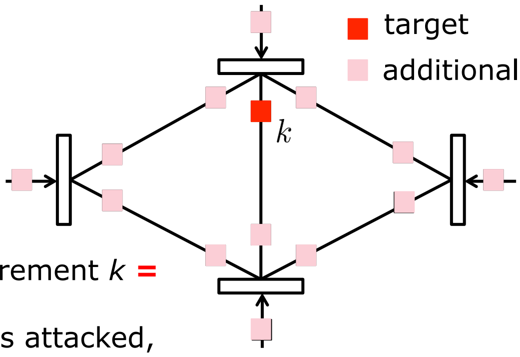
Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

A Security Index

Stealth attack $\Delta y = H\Delta\theta$

In general, $e_k \notin \text{span}(H)$



Security index for measurement *k* =

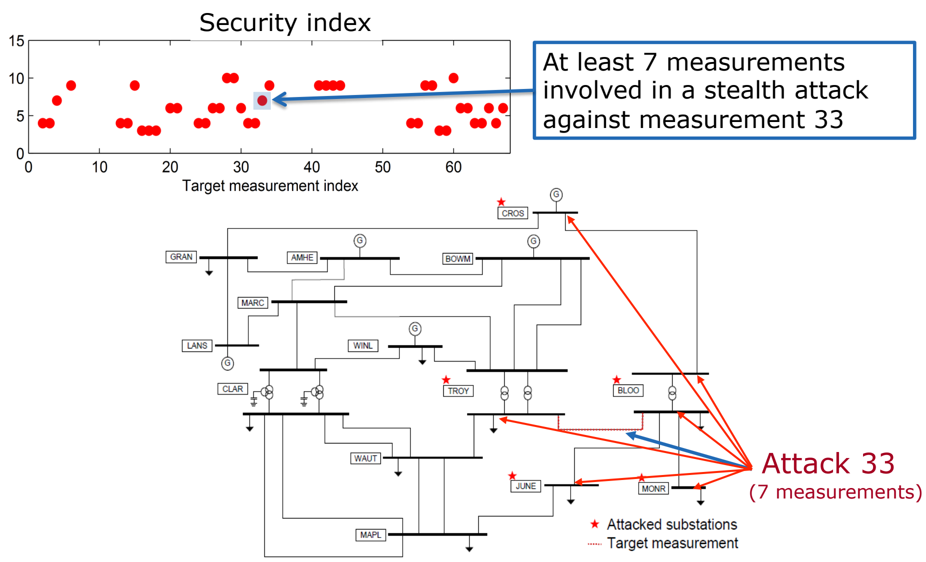
Minimum number of meters attacked, targeting the k^{th} measurement:

$$\min_{\Delta\theta} \|H\Delta\theta\|_0$$
$$\text{s.t.} \;\; H(k,:)\Delta\theta = 1$$

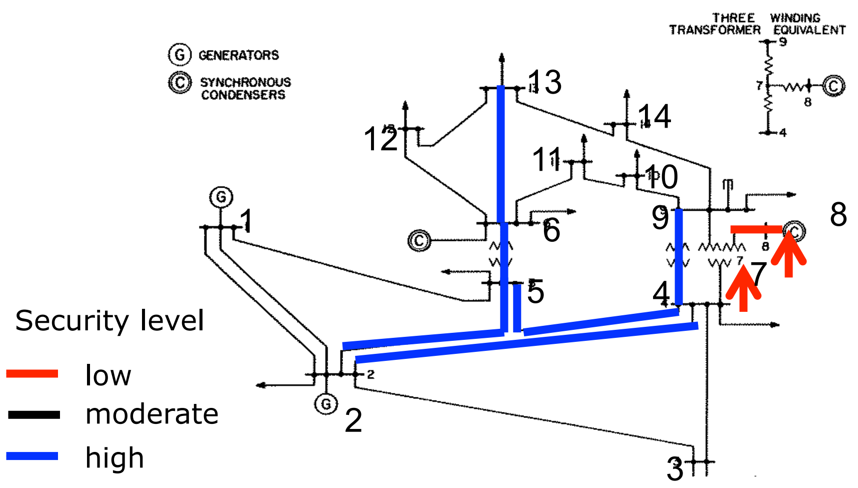
38

[Sandberg *et al*., CPSWEEK, 2010; Kosut *et al*., IEEE TSG, 2011]

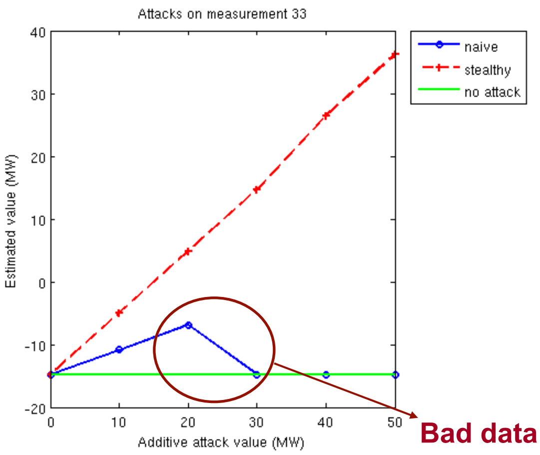
Security Indices for 40-bus Network



Quantify Security to Aid Allocation of Protection



Verification on SCADA Testbed



False value (MW)	Estimated value (MW)	# BDD Alarms
-14.8	-14.8	0
35.2	36.2	0
85.2	86.7	0
135.2	137.5	0
185.2	Non convergent	-

Bad data detected & removed

- Stealth attack of 150 MW (55% of nominal value) passed undetected in testbed!

[Teixeira *et al.*, IFAC WC, 2011]

How Hard is it to Compute the Security Index?

$$\min_{\Delta\theta} \|H\Delta\theta\|_0$$
$$\text{s.t.}\ \ H(k,:)\Delta\theta = 1$$

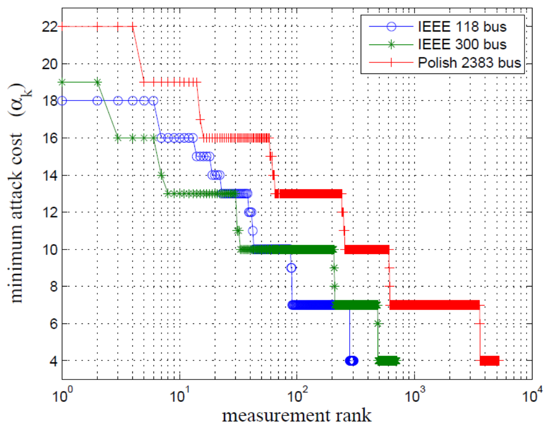
Problem known to be **NP-hard** for arbitrary *H*, but it is possible to explore structure

Method/Example	118 bus	300 bus	2383 bus
MILP	763 sec	6708 sec	About 5.7 days
Min Cut	0.3 sec	1 sec	31 sec

42

[Sou et al., IEEE TSG, 2014; Hendrickx et al., IEEE TAC, 2014]

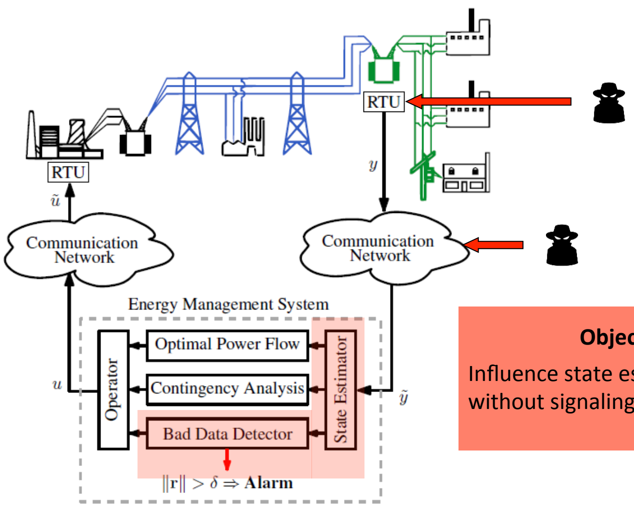
Large-Scale Examples



Method/Example	118 bus	300 bus	2383 bus
MILP	763 sec	6708 sec	About 5.7 days
Min Cut	0.3 sec	1 sec	31 sec

[Sou et al., IEEE TSG, 2014; Hendrickx et al., IEEE TAC, 2014]

Attack Scenario (so far)



Objective

Influence state estimator without signaling bad data alarm

[Giani *et al.*, IEEE ISRCS, 2009; Mohajerin Esfahani *et al.*, CDC, 2010]

Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

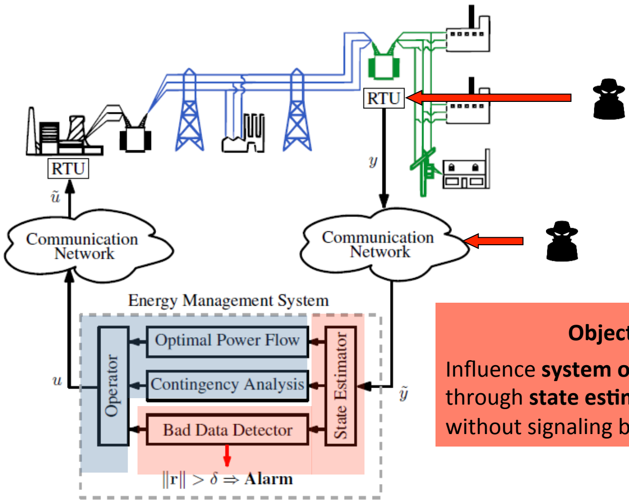
Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data

  - Static systems
  - Dynamic systems

- Conclusions

Closing the loop over corrupted data

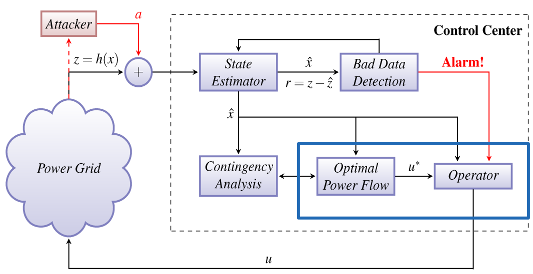


Objective

Influence **system operation** through **state estimator + controller** without signaling bad data alarm

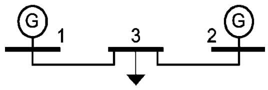
[Static case: Teixeira *et al.*, ACC, 2013; Dynamic case: Teixeira, PhD Thesis, 2014]

Cyber Security of Optimal Power Flow



- How do stealthy attacks **affect the power system's operation**?
  - Related work: [Xie et al, 2010], [Yuan et al, 2011]
- Optimal Power Flow
  - Computes generator set points minimizing operation costs
  - Ensures operation constraints

DC-Optimal Power Flow



- DC-Optimal Power Flow considers the lossless DC model
  - $P^d \in \mathbb{R}^N$ power demand
  - $P^g \in \mathbb{R}^{N_g}$ power generation
- Operation costs:

  $c(P^g) = \frac{1}{2} P^{g\top} Q P^g + R^\top P^g + C_0$
  - Generation costs
  - Transmission losses
- Optimal power generation

$$\begin{aligned} \min_{P^g} \quad & c(P^g) \\ \text{s.t.} \quad & g(P^g, P^d) = \mathbf{1}^\top P^g + \mathbf{1}^\top P^d = 0 \\ & f(P^g, P^d) = F_g P^g + F_d P^d + F_0 \leq 0 \end{aligned}$$

DC-Optimal Power Flow

- Lagrangian function:

$$L(P^g, \nu, \lambda) = c(P^g) + \nu(\mathbf{1}^\top P^g + \mathbf{1}^\top P^d) + \lambda^\top (F_g P^g + F_d P^d + F_0)$$

- At optimality, the KKT conditions hold:

$$\underbrace{\begin{bmatrix} Q & F_g^\top & \mathbf{1} \\ \mathbf{1}^\top & 0 & 0 \\ H_1 F_g & 0 & 0 \\ 0 & H_0 & 0 \end{bmatrix}}_{K} \begin{bmatrix} P^{g*} \\ \lambda^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -R \\ -\mathbf{1}^\top P^d \\ H_1(-F_d P^d - F_0) \\ 0 \end{bmatrix}$$

DC-Optimal Power Flow under Attack

- The estimate $\hat{P}^d$ is given by the **State Estimator**
  - vulnerable to cyber attacks
- Suppose the system is in optimality with $\hat{P}^d = P^d$ and $\hat{P}^g = P^{g*}$
- Operation under Data Attack

$$\hat{P}^g_a = P^{g*} + a_g \qquad \hat{P}^d_a = P^d + a_d$$

Ficticious operating conditions

$$\Rightarrow \quad \min_{P^g} \; c(P^g) \quad \text{s.t.} \quad g(P^g, \hat{P}^d_a) = 0, \quad f(P^g, \hat{P}^d_a) \le 0 \quad \Rightarrow \quad \hat{P}^{g*}_a$$

Proposed control action

- When would an operator apply the proposed control action?
- What would be the resulting operating cost?

DC-Optimal Power Flow under Attack

- Assume the attack does not change the active constraints
  - thus H_1, H_0 are known
- The proposed control action is given by

$$\begin{bmatrix} \hat{P}^{g*}_a - P^{g*} \\ \hat{\lambda}^*_a - \lambda^* \\ \hat{\nu}^*_a - \nu^* \end{bmatrix} = K^{-1} \begin{bmatrix} 0 \\ -\mathbf{1}^\top \\ -H_1 F_d \\ 0 \end{bmatrix} a_d = \begin{bmatrix} T_g \\ T_\lambda \\ T_\nu \end{bmatrix} a_d,$$

- $\hat{P}^{g*}_a$ is an affine map w.r.t a_d

Estimated Re-Dispatch Profit

$\hat{P}^g_a = P^{g*} + a_g$
$\hat{P}^d_a = P^d + a_d$
Fictitious operating conditions

$$\min_{P^g} \; c(P^g) \quad \text{s.t.} \quad g(P^g, \hat{P}^d_a) = 0, \; f(P^g, \hat{P}^d_a) \le 0$$

$\hat{P}^{g*}_a$
Proposed control action

- Consider the corrupted estimates $\hat{P}^d_a$ and $\hat{P}^g_a$
  - $c(\hat{P}^g_a)$: estimated operation cost
  - $c(\hat{P}^{g*}_a)$: estimated optimal operation cost given $\hat{P}^d_a$
  - $\hat{\mathcal{P}}_a \triangleq c(\hat{P}^g_a) - c(\hat{P}^{g*}_a)$: **estimated re-dispatch profit**
- Large estimated profit may lead the operator to apply $\hat{P}^{g*}_a$

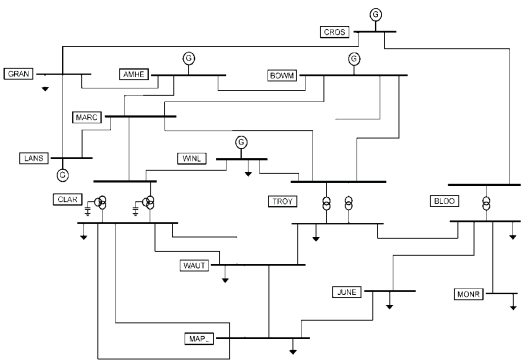
True Re-Dispatch Profit

$\hat{P}^{g*}_a$ → Slack generators → P^g_a

Proposed control action — True generation profile

- Mismatches between $\hat{P}^d_a$ and P^d are compensated by slack generators
  - can be modeled as an affine map w.r.t a_d : $P^{g*}_a - P^{g*} = MT_g a_d$
  - $c(P^g_a)$: true operation cost after re-dispatch
  - $\mathcal{P}_a \triangleq c(P^{g*}) - c(P^{g*}_a)$: **true re-dispatch profit**
- Large $|\mathcal{P}_a|$ corresponds to attacks with higher impact

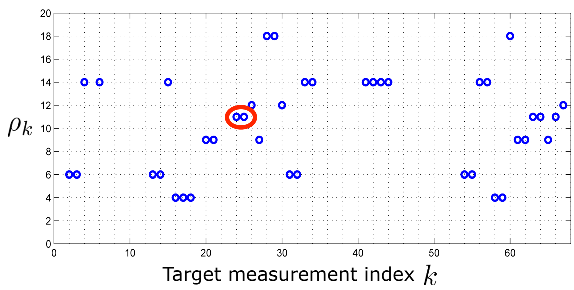
VIKING Benchmark: Impact of Data Attacks



- Cost function corresponds to the total resistive losses
- Sparse attacks are computed based on the security metric
- $\mathcal{P}_a$ is computed for each sparse attack

VIKING Benchmark: Impact of Data Attacks

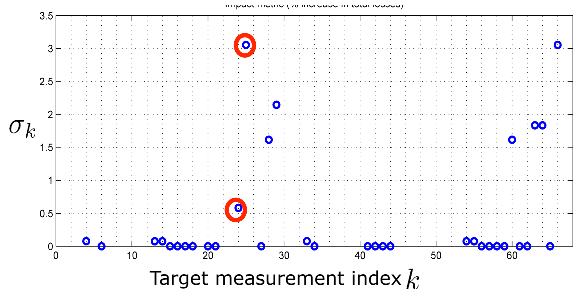
- Security metric $\rho_k = \|a^*\|_0$
  - Do all sparse attacks have equal impact?



- Impact of Data Attacks

$$\frac{\mathcal{P}_a^*}{c(P^{g*})}$$

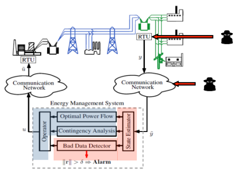
  - Most sparse attacks have low impact on operation cost

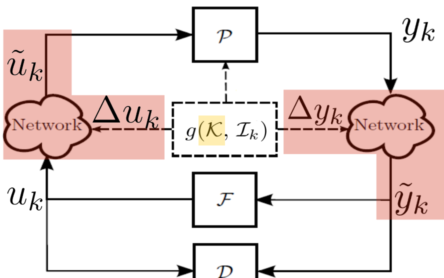


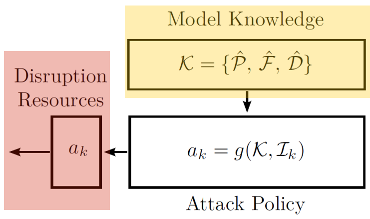
Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
  - Static systems
  - Dynamic systems
- Conclusions

Adversary Model



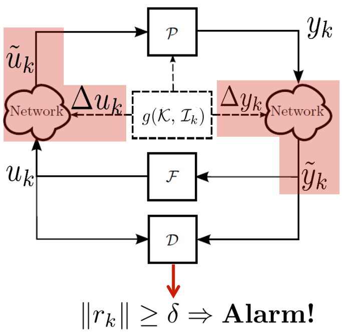




- **Attack policy:** Maximize impact on plant's state without alarms
- **Model knowledge:** Dynamical model of the closed-loop system
- **Disruption resources:** Small no. of measurement and actuation channels

How resilient is the system against such an attack?

Stealthy Additive Deception Attack



- Closed-loop system under attack:

$$x_{k+1} = Ax_k + Ba_k \qquad a_k = \begin{bmatrix}\Delta u_k \\ \Delta y_k\end{bmatrix}$$
$$r_k = Cx_k + Da_k$$

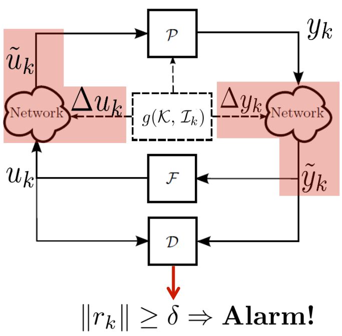
- Stealthy attack:

  Input sequence that attains a zero output r_k

$$\{a_k\}_{k=0}^{\infty} : \; r_k \approx 0, \;\; \forall k$$

  Can be derived from the system's zero dynamics

Maximum-Impact Stealthy Attack



- Closed-loop system under attack:

$$x_{k+1} = Ax_k + Ba_k \qquad a_k = \begin{bmatrix}\Delta u_k \\ \Delta y_k\end{bmatrix}$$
$$r_k = Cx_k + Da_k$$

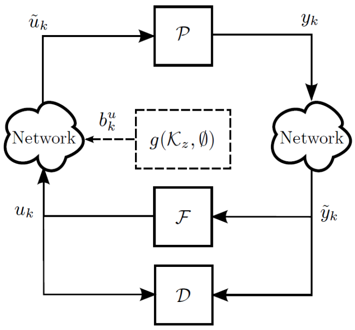
- Maximum-impact stealthy attack:
  - Maximize "energy" of the state signal
  - Keep the output signal "small"

$$\begin{aligned} \underset{\{a_k\}_{k=0}^{\infty}}{\text{maximize}} \quad & \textstyle\sum_{k=0}^{\infty} \|x_k\|_2^2 \\ \text{subject to} \quad & \textstyle\sum_{k=0}^{\infty} \|r_k\|_2^2 \leq \delta \end{aligned}$$

- If the system has **unstable zero-dynamics**:
  - There exists an *exponentially increasing* input that attains a "small" output

$$\{a_k\}_{k=0}^{\infty} : \; r_k \approx 0, \;\; \forall k$$
$$\|a_k\| \to \infty, \quad \|x_k\| \to \infty$$

Zero Dynamics Attack



- Zero dynamics are characterized by:

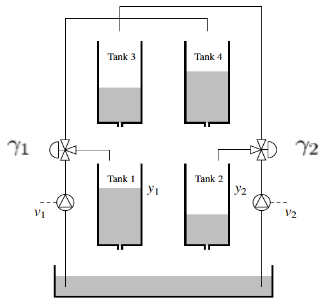
$$\begin{bmatrix} \nu I - A & -B \\ C & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ g \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

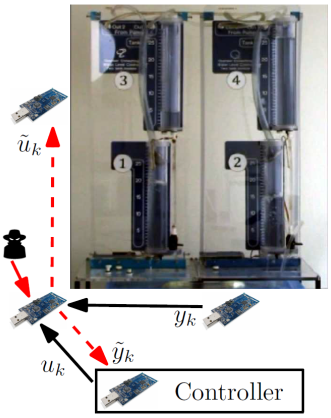
- Suggests attack on actuators with policy:

$$a_k = g\nu^k$$

- If the zero is unstable, then the plant state can be made arbitrarily large by this attack without detection
- Requires system knowledge (zero dynamics) but no disclosure resources

Experimental Set-Up





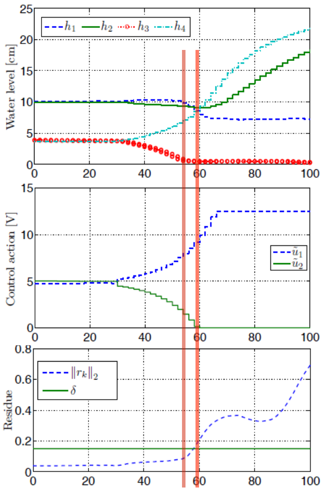
$$\frac{dx}{dt} = \begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{A_3}{A_1 T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{A_4}{A_2 T_4} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix} x + \begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix} u$$

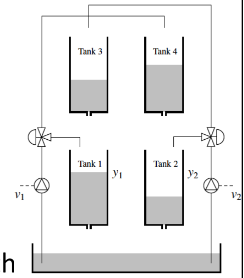
$$y = \begin{bmatrix} k_c & 0 & 0 & 0 \\ 0 & k_c & 0 & 0 \end{bmatrix} x$$

Quadruple-tank process has unstable zero dynamics if $0 < \gamma_1 + \gamma_2 < 1$

[J, 2000]

Experimental Validation





- **Attack goal:** Empty Tank 3
- Zero dynamics attack on both actuators starts at t=30s
- Tank 3 becomes empty at t=55s
- The attack is detected at t=58s
- Actuator 2 saturates at t=60s

Teixeira et al, Automatica, 2015

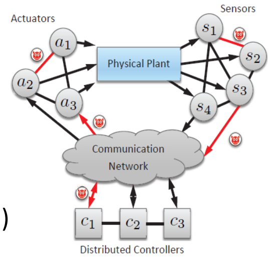
Outline

- Introduction
- Adversary model for networked control systems
- Attacks on power network state estimator
- Security index for stealthy minimum-effort attacks
- Closing the loop over corrupted data
- Conclusions

Research Program in Cyber-Physical Security

Need **analysis and design tools** to understand and mitigate attacks

- Which threats should we care about?
- Which resources are more important to protect?
- What impact can we expect of an attack?
- How to create resilient systems?



Cross-disciplinary research agenda

- IT security (authentication, encryption, firewalls, etc.) is needed, but not sufficient
- Malicious actions can enter in the control loop, even if channels are secure

Grand societal challenges

- Impact on future infrastructure systems where everything is connected
- Systems need to be trusted by the general public

Conclusions

- **Cyber-security models** for networked control systems
- Undetectable **false-data attacks** against state estimator, both in theory and practice
- **Security index** to estimate vulnerabilities
- Suggests locations of counter measures
- **Further studies** needed on integrating cyber and physical security with social and human behaviors

SPARKS SMART GRID PROTECTION AGAINST CYBER ATTACKS

https://project-sparks.eu/

http://people.kth.se/~kallej