Security in CPS

Paulo Tabuada

Cyber-Physical Systems Laboratory
Department of Electrical Engineering
University of California at Los Angeles

CyPhyLab UCLA

Control systems' security in the news

PCWorld
Business Center

Discover news, guides, and products for your business

Software & Services | Office Hardware | Security | Servers & Storage | Cell Phones & Mobile | C

Sign in with or Create a New Account.

Recommend: Like 957 179 +1 1 1K Email 5 Comments Print

BUSINESS CENTER Sep 21, 2010 4:10 am

Was Stuxnet Built to Attack Iran's Nuclear Program?

By Robert McMillan, IDG News

A highly sophisticated computer worm that has spread through Iran, Indonesia and India was built to destroy operations at one target: possibly Iran's Bushehr nuclear reactor.

SIMILAR ARTICLES:

Duqu: New Malware Is Stuxnet 2.0

Stuxnet Compromise at Iranian Nuclear Plant May Be By Design

That's the emerging consensus of security experts who have examined the Stuxnet worm. In recent weeks, they've broken the cryptographic code behind the software and taken a look at how the worm operates in test environments. Researchers studying the worm all agree that Stuxnet was

CyPhyLab UCLA

Control systems' security in the news

PCWorld
Business Center
Discover news, guides, and products for your

ArmyTimes A GANNETT COMPANY

HOME | NEWS | BENEFITS | MONEY | CAREERS & EDUCATION | COMMUNITY
MARKETPLACE | CLASSIFIEDS

Army News
All Army News
Guard & Reserve
This Week's Issue
Subscribe to RSS

Quick Links
Hall of Valor
Army Discussions
Frontline Photos

Like 2 Tweet 14 SHARE

Cyber attacks on utilities, industries rise

By Douglas Birch - The Associated Press
Posted : Thursday Sep 29, 2011 19:50:39 EDT

IDAHO FALLS, Idaho — U.S. utilities and other crucial industries face an increasing number of cyber break-ins by attackers using more sophisticated methods, a senior Homeland Security Department official told reporters during the first tour of the government's secretive defense labs intended to protect the nation's power grid, water and communications systems.

CyPhyLab UCLA

Control systems' security in the news

Technology | DOI:10.1145/2018396.2018403

PCWorld
Busine
Army
HOME
MARKE
Army
All Army
Guard &
This Weel
Subs
Quick Link
Hall of Va
Army Disc

Hacking Cars

Researchers have discovered important security flaws in modern automobile systems. Will car thieves learn to pick locks with their laptops?

NOT SO LONG ago, car thieves plied their trade with little more than a coat hanger and a screwdriver. New anti-theft technologies have made today's cars much harder to steal, but the growing tangle of computer equipment under the modern hood is creating new security risks that carmakers are just beginning to understand.

Ever since Toyota's well-publicized struggles with the computerized braking systems in its 2010 Prius hybrid cars, automotive computer systems have come under increasing scrutiny. In the last few years, researchers have identified a range of new, unexpected security flaws that could potentially affect large numbers of new cars. Given the specialized programming knowl-

PhyLab UCLA

Control systems' security in the news

Technology | DOI:10.1145/2018396.2018403

PCWorld

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIGN OPIN

THREAT LEVEL | german steel mill | icymi | Sony | stuxnet

A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever

BY KIM ZETTER 01.08.15 | 5:30 AM | PERMALINK

understand.

Ever since Toyota's well-publicized struggles with the computerized braking systems in its 2010 Prius hybrid cars, automotive computer systems have come under increasing scrutiny. In the last few years, researchers have identified a range of new, unexpected security flaws that could potentially affect large numbers of new cars. Given the specialized programming knowl-

PhyLab UCLA

Control systems' security in the news

Technology | DOI:10.1145/2018396.2018403

PCWorld

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIGN OPIN

THREAT LEVEL | german steel mill | icymi | Sony | stuxnet

A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever

I'm referring to the revelation, in a German report released just before Christmas (.pdf), that hackers had struck an unnamed steel mill in Germany. They did so by manipulating and disrupting control systems to such a degree that a blast furnace could not be properly shut down, resulting in "massive"—though unspecified—damage.

have come under increasing scrutiny. In the last few years, researchers have identified a range of new, unexpected security flaws that could potentially affect large numbers of new cars. Given the specialized programming knowl-

PhyLab UCLA

Control systems' security in the news

Technology | DOI:10.1145/2018396.2018403

PCWorld

WIRED GEAR SCIENCE ENTERTAINMENT BUSINESS SECURITY DESIGN OPIN

THREAT LEVEL | german steel mill | icymi | Sony | stuxnet

A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever

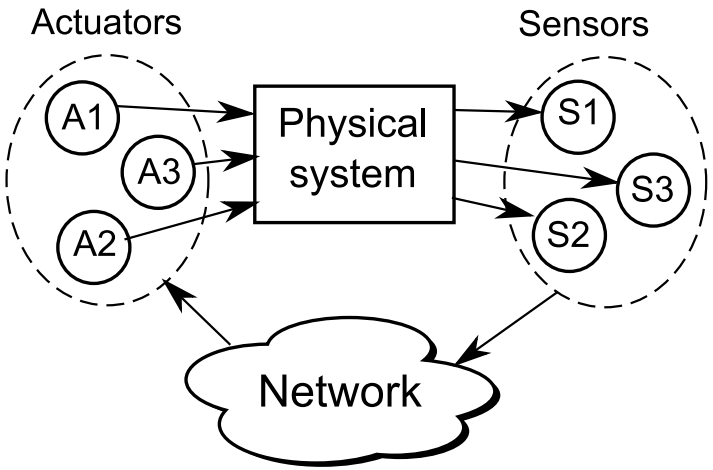
I'm referring to the revelation, in a German report released just before Christmas (.pdf), that hackers had struck an unnamed steel mill in Germany. They did so by manipulating and disrupting control systems to such a degree that a blast furnace could not be properly shut down, resulting in "massive"—though unspecified—damage.

have come under increasing scrutiny. In the last few years, researchers have identified a range of new, unexpected security flaws that could potentially affect large numbers of new cars. Given the specialized programming knowl-

PhyLab UCLA

Why security for control systems?

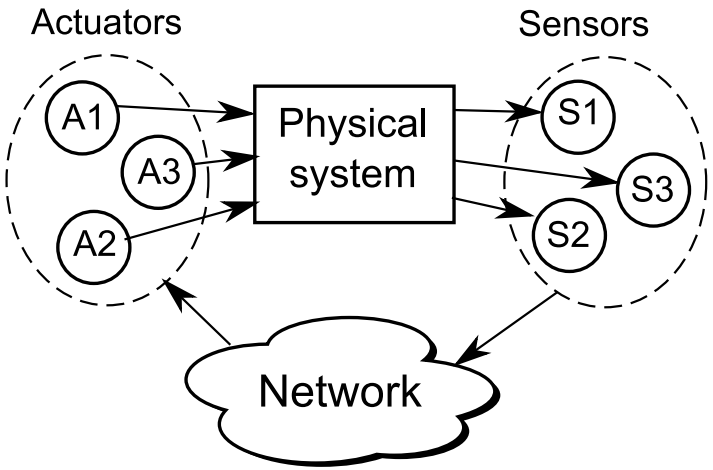
- Control systems are physical processes (chemical plants, power grids, aircraft, etc.) controlled by cyber components (computation and communication hardware/software).



- Control systems are becoming increasingly larger, distributed, and open to the cyber-world (e.g., internet): **increased vulnerability to attacks.**

Why security for control systems?

- Control systems are physical processes (chemical plants, power grids, aircraft, etc.) controlled by cyber components (computation and communication hardware/software).

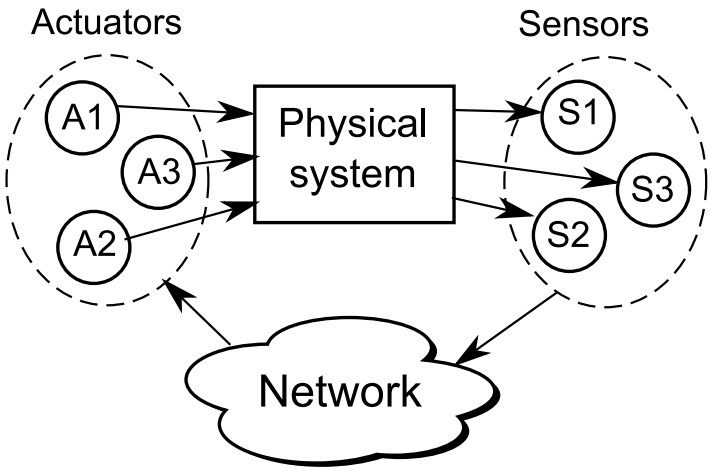


- Control systems are becoming increasingly larger, distributed, and open to the cyber-world (e.g., internet): **increased vulnerability to attacks.**
- Examples of real attacks: Sewage system (Australia, 2000), Natural gas pipelines (Russia, 2000), Stuxnet (Iran, 2010), Steel mill (Germany, 2014), ...

CyPhyLab UCLA

Why security for control systems?

- Control systems are physical processes (chemical plants, power grids, aircraft, etc.) controlled by cyber components (computation and communication hardware/software).



- Control systems are becoming increasingly larger, distributed, and open to the cyber-world (e.g., internet): **increased vulnerability to attacks.**
- Examples of real attacks: Sewage system (Australia, 2000), Natural gas pipelines (Russia, 2000), Stuxnet (Iran, 2010), Steel mill (Germany, 2014), ...
- **We need efficient ways to cope with attacks on control systems.**

CyPhyLab UCLA

Security for control systems

- Fault-tolerant control and robust control deal with disturbances to the physical process and not with adversarial attacks:
  - Fault-tolerant control: fixed number of failure modes;
  - Robust control: bounded disturbances or known statistical model.

Security for control systems

- Fault-tolerant control and robust control deal with disturbances to the physical process and not with adversarial attacks:
  - Fault-tolerant control: fixed number of failure modes;
  - Robust control: bounded disturbances or known statistical model.
- This talk: how to control a linear system when the sensors are under attack?

CyPhyLab UCLA

Security for control systems

- Fault-tolerant control and robust control deal with disturbances to the physical process and not with adversarial attacks:
  - Fault-tolerant control: fixed number of failure modes;
  - Robust control: bounded disturbances or known statistical model.
- This talk: how to control a linear system when the sensors are under attack?
- Other problems investigated in the literature:
  - Security risk management for CPS
  - Cross-layer security via game theory
  - Energy theft
  - Covert misappropriation of plants
  - Physical watermarking
  - Detectability and identifiability of attacks
- Entry point: special issue on CPS Security, Control Systems Magazine, 35(1), 2015.

CyPhyLab UCLA

Lecture plan

- Control under sensor attacks
- Attacking sensors
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- State reconstruction under sensor attacks:
  - Convex relaxation of an ℓ_0 optimization problem
  - Observer-based state reconstruction
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

Lecture plan

- **Control under sensor attacks**
- Attacking sensors
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- State reconstruction under sensor attacks:
  - Convex relaxation of an ℓ_0 optimization problem
  - Observer-based state reconstruction
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t)$$

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- Some sensors are attacked:
  - $e_i(t) \neq 0 \longrightarrow$ sensor i is attacked at time t;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- Some sensors are attacked:
  - $e_i(t) \neq 0 \longrightarrow$ sensor i is attacked at time t;
  - If sensor i is attacked, $e_i(t)$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.);

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- Some sensors are attacked:
  - $e_i(t) \neq 0 \longrightarrow$ sensor i is attacked at time t;
  - If sensor i is attacked, $e_i(t)$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.);
- Set of attacked sensors (unknown) is denoted by $K \subset \{1, \ldots, p\}$:

$$\text{supp}(e) = \{i \in \{1, \ldots, p\} \mid e_i \neq 0\} = K$$

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- Some sensors are attacked:
  - $e_i(t) \neq 0 \longrightarrow$ sensor i is attacked at time t;
  - If sensor i is attacked, $e_i(t)$ can be **arbitrary** (no boundedness assumption, no stochastic model, etc.);
- Set of attacked sensors (unknown) is denoted by $K \subset \{1, \ldots, p\}$:

$$\text{supp}(e) = \{i \in \{1, \ldots, p\} \mid e_i \neq 0\} = K$$

- Number of attacked sensors will be denoted by q: $q = |K|$;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- S

$$\big[e(0)|e(1)|e(2)|e(3)\big] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 3 & 1 & -5 & 3 \\ -5 & 3 & -2 & -5 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad K = \{2,3\}, \quad |K| = 2$$

otion,

- Set of attacked sensors (unknown) is denoted by $K \subset \{1,\dots,p\}$:

$$\text{supp}(e) = \{i \in \{1,\dots,p\} \mid e_i \neq 0\} = K$$

- Number of attacked sensors will be denoted by q: $q = |K|$;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- The objective is to design a controller:

$$u(t) = \phi_\alpha(t, y(0), \ldots, y(t))$$

rendering the closed-loop system exponentially stable:

$$\|x(t)\| \leq \kappa \alpha^t \|x(0)\| \qquad \forall x(0) \in \mathbb{R}^n$$

for any chosen rate of decay $\alpha \in [0, 1[$

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- The objective is to design a controller:

$$u(t) = \phi_\alpha(t, y(0), \ldots, y(t))$$

rendering the closed-loop system exponentially stable:

$$\|x(t)\| \leq \kappa \alpha^t \|x(0)\| \qquad \forall x(0) \in \mathbb{R}^n$$

for any chosen rate of decay $\alpha \in [0, 1[$, notwithstanding any adversarial attack to q sensors, i.e., for all $e(t) \in \mathbb{R}^p$ with support K, $|K| = q$.

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- We can see this as a game between the controller and the attacker:

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- We can see this as a game between the controller and the attacker:
  - The matrices A, B, and C are known to the controller but $x(0)$ is not. The attacker is omniscient;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- We can see this as a game between the controller and the attacker:
  - The matrices A, B, and C are known to the controller but $x(0)$ is not. The attacker is omniscient;
  - The controller chooses an action $u(t) \in \mathbb{R}^m$ at time $t \in \mathbb{N}_0$ based on all its past observations $y(0), y(1), \ldots, y(t)$;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- We can see this as a game between the controller and the attacker:
  - The matrices A, B, and C are known to the controller but $x(0)$ is not. The attacker is omniscient;
  - The controller chooses an action $u(t) \in \mathbb{R}^m$ at time $t \in \mathbb{N}_0$ based on all its past observations $y(0), y(1), \ldots, y(t)$;
  - The attacker chooses K before the game starts and during the game it chooses $e(t) \in \mathbb{R}^p$ with support in K at time $t \in \mathbb{N}_0$;

CyPhyLab UCLA

The setup

- Physical process modeled as a linear dynamical system:

$$x(t+1) = Ax(t) + Bu(t)$$

- A total of p sensors monitor state of plant ($y(t) \in \mathbb{R}^p$):

$$y(t) = Cx(t) + \underbrace{e(t)}_{\substack{\text{attack} \\ \text{vector}}}$$

- We can see this as a game between the controller and the attacker:
  - The matrices A, B, and C are known to the controller but $x(0)$ is not. The attacker is omniscient;
  - The controller chooses an action $u(t) \in \mathbb{R}^m$ at time $t \in \mathbb{N}_0$ based on all its past observations $y(0), y(1), \ldots, y(t)$;
  - The attacker chooses K before the game starts and during the game it chooses $e(t) \in \mathbb{R}^p$ with support in K at time $t \in \mathbb{N}_0$;
  - The controller seeks to enforce the inequality $\|x(t)\| \leq \kappa\alpha^t\|x(0)\|$ while the attacker seeks to prevent it.

CyPhyLab UCLA

Questioning the setup

- Why is the set K of attacked sensors fixed throughout the game?

Questioning the setup

- Why is the set K of attacked sensors fixed throughout the game?
  - Compromising a sensor takes time. While the attacker is working to compromise one additional sensor we can treat K as fixed.

Questioning the setup

- Why is the set K of attacked sensors fixed throughout the game?
  - Compromising a sensor takes time. While the attacker is working to compromise one additional sensor we can treat K as fixed.
- Is the attacker attacking the sensors or the communication between the sensors and the controller?

Questioning the setup

- Why is the set K of attacked sensors fixed throughout the game?
  - Compromising a sensor takes time. While the attacker is working to compromise one additional sensor we can treat K as fixed.
- Is the attacker attacking the sensors or the communication between the sensors and the controller?
  - In our mathematical setup there is no need to distinguish between these two cases.

CyPhyLab UCLA

Questioning the setup

- Why is the set K of attacked sensors fixed throughout the game?
  - Compromising a sensor takes time. While the attacker is working to compromise one additional sensor we can treat K as fixed.
- Is the attacker attacking the sensors or the communication between the sensors and the controller?
  - In our mathematical setup there is no need to distinguish between these two cases.
- Can you not protect the sensors or the communication using cyber-security techniques?

CyPhyLab UCLA

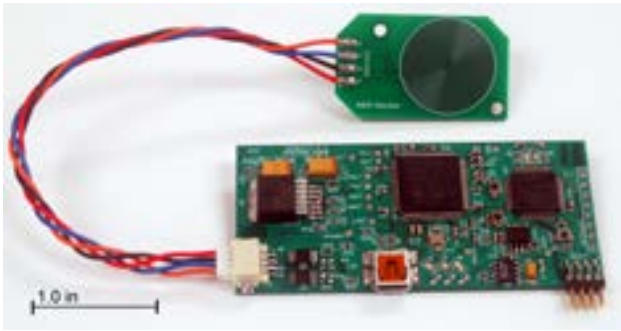
Lecture plan

- Control under sensor attacks
- **Attacking sensors**
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- State reconstruction under sensor attacks:
  - Convex relaxation of an ℓ_0 optimization problem
  - Observer-based state reconstruction
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

Attacking sensors

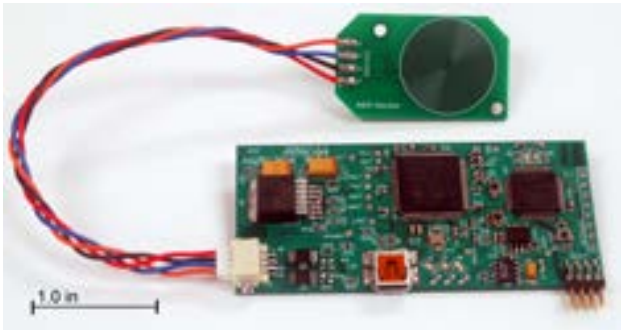
Attacking sensors

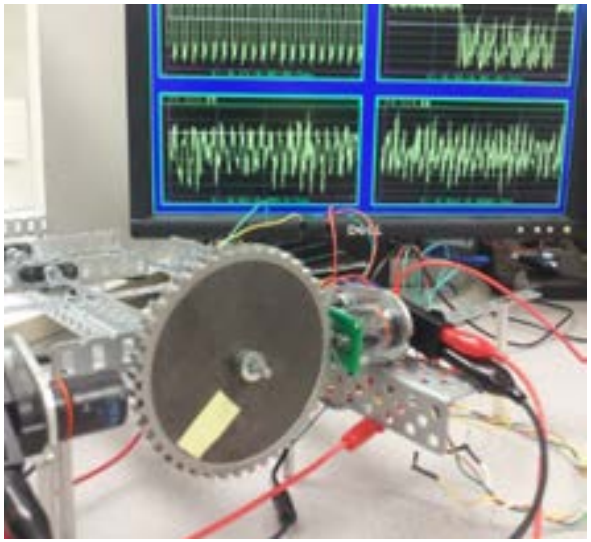


Noninvasive spoofing attacks for Anti-Lock Braking systems
Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava.
Workshop on Cryptographic Hardware and Embedded Systems, 2013 (CHES 2013).

CyPhyLab UCLA

Attacking sensors

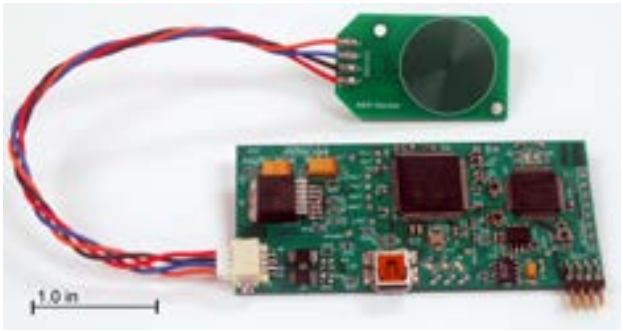


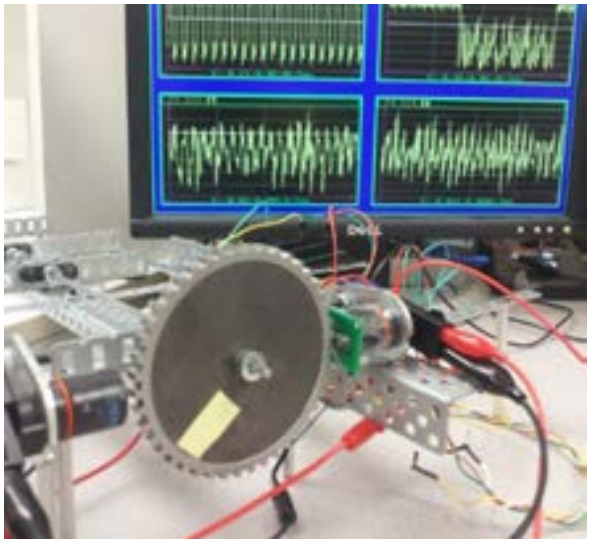


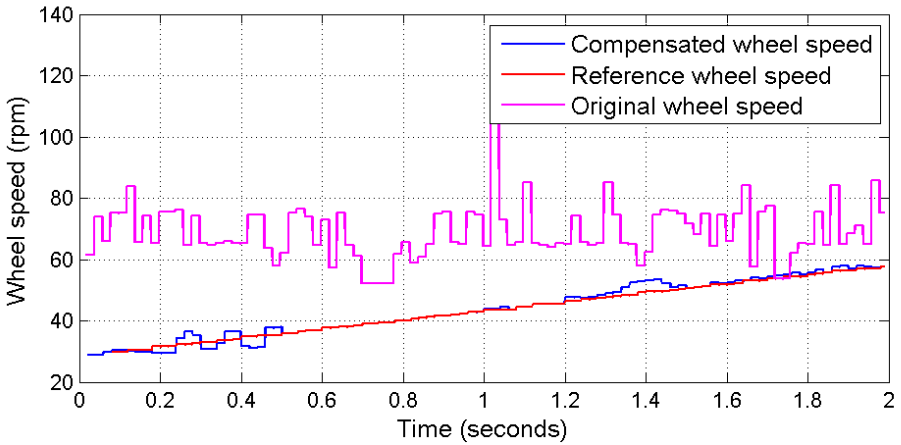
Noninvasive spoofing attacks for Anti-Lock Braking systems
Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava.
Workshop on Cryptographic Hardware and Embedded Systems, 2013 (CHES 2013).

CyPhyLab UCLA

Attacking sensors



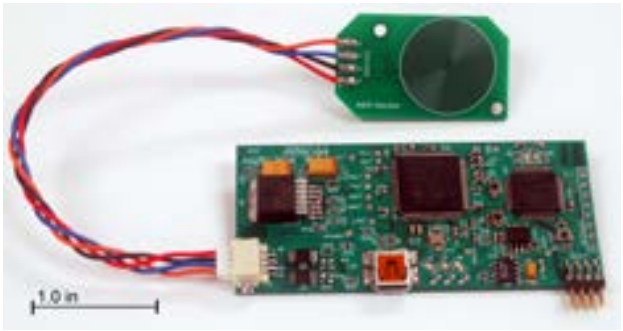


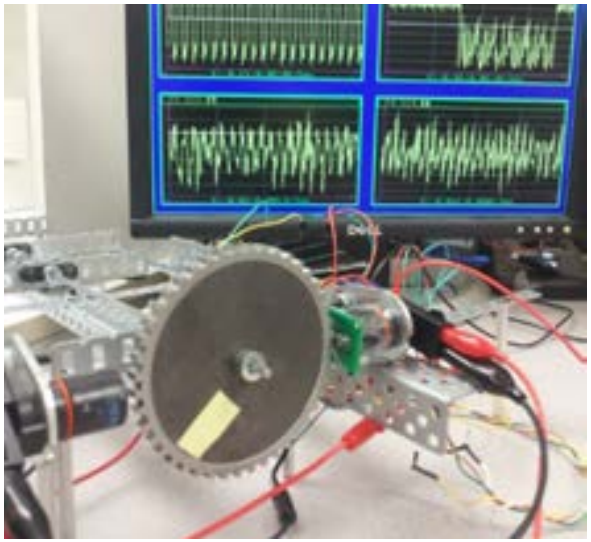


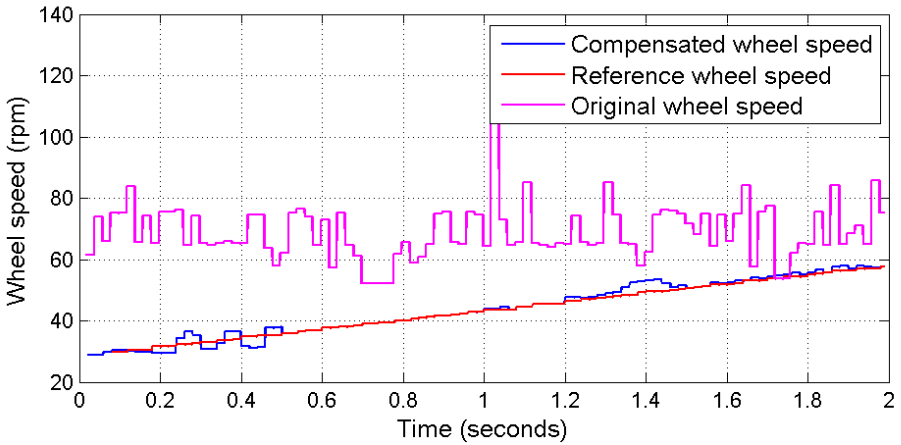
Noninvasive spoofing attacks for Anti-Lock Braking systems
Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava.
Workshop on Cryptographic Hardware and Embedded Systems, 2013 (CHES 2013).

CyPhyLab UCLA

Attacking sensors







Noninvasive spoofing attacks for Anti-Lock Braking systems
Y. Shoukry, P. Martin, P. Tabuada, and M. Srivastava.
Workshop on Cryptographic Hardware and Embedded Systems, 2013 (CHES 2013).

CyPhyLab UCLA

Lecture plan

- Control under sensor attacks
- Attacking sensors
- **Some fundamental results**
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- State reconstruction under sensor attacks:
  - Convex relaxation of an ℓ_0 optimization problem
  - Observer-based state reconstruction
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

A separation result

- The attacks are arbitrary, in particular they can depend on the state in a nonlinear and time-varying manner.
- Do we need to design a nonlinear, time-varying, and dynamic controller resilient to attacks?

A separation result

- The attacks are arbitrary, in particular they can depend on the state in a nonlinear and time-varying manner.
- Do we need to design a nonlinear, time-varying, and dynamic controller resilient to attacks?

Theorem

Consider the linear control system:

$$x(t+1) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + e(t).$$

If there exists a controller $u(t) = \phi(t, y(0), \ldots, y(t))$ *rendering the closed-loop system exponentially stable:*

$$\|x(t)\| \leq \kappa \alpha^t \|x(0)\| \qquad \forall x(0) \in \mathbb{R}^n$$

for a sufficiently fast rate of decay ($\alpha < \min_i |\lambda_i(A)|$) *and despite an adversarial attack to q sensors, i.e., for all* $e(t) \in \mathbb{R}^p$ *with support* K, $|K| = q$, *then there exists a decoder* $D : \mathbb{R}^{n \times p} \to \mathbb{R}^n$ *that correctly reconstructs the state in n steps:*

$$x(t-n+1) = D(y(t-n+1), \ldots, y(t)).$$

A separation result

Theorem

Consider the linear control system:

$$x(t+1) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + e(t).$$

If there exists a controller $u(t) = \phi(t, y(0), \ldots, y(t))$ rendering the closed-loop system exponentially stable:

$$\|x(t)\| \leq \kappa\alpha^t\|x(0)\| \qquad \forall x(0) \in \mathbb{R}^n$$

for a sufficiently fast rate of decay ($\alpha < \min_i |\lambda_i(A)|$) and despite an adversarial attack to q sensors, i.e., for all $e(t) \in \mathbb{R}^p$ with support K, $|K| = q$, then there exists a decoder $D : \mathbb{R}^{n\times p} \to \mathbb{R}^n$ that correctly reconstructs the state in n steps:

$$x(t-n+1) = D(y(t-n+1), \ldots, y(t)).$$

We can solve our initial problem in two steps:

1. design the decoder D;
2. design a linear static controller.

CyPhyLab UCLA

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;
- A **decoder** D processes observations $y(0), \ldots, y(T-1)$ and produces an estimate of the initial state $x(0)$.

CyPhyLab UCLA

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;
- A **decoder** D processes observations $y(0), \ldots, y(T-1)$ and produces an estimate of the initial state $x(0)$.
- We say that a decoder $D : (\mathbb{R}^p)^T \to \mathbb{R}^n$ **corrects q errors after T steps** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x(0) \in \mathbb{R}^n$, and for any attack vectors $e(0), \ldots, e(T-1)$ with support K, $|K| = q$, we have:

$$D(y(0), \ldots, y(T-1)) = x(0).$$

CyPhyLab UCLA

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;
- A **decoder** D processes observations $y(0), \ldots, y(T-1)$ and produces an estimate of the initial state $x(0)$.
- We say that a decoder $D : (\mathbb{R}^p)^T \to \mathbb{R}^n$ **corrects q errors after T steps** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x(0) \in \mathbb{R}^n$, and for any attack vectors $e(0), \ldots, e(T-1)$ with support K, $|K| = q$, we have:

$$D(y(0), \ldots, y(T-1)) = x(0).$$

- We say that q **errors are correctable**, for the system (A, C), if there exists a decoder that can correct q errors.

CyPhyLab UCLA

Error correction

$$\begin{aligned} x(t+1) &= Ax(t) \\ y(t) &= Cx(t) + e(t) \end{aligned}$$

- We assume the input to be known since we design the controller. For simplicity we will take $u(t) = 0$ for all $t \in \mathbb{N}$;
- A **decoder** D processes observations $y(0), \ldots, y(T-1)$ and produces an estimate of the initial state $x(0)$.
- We say that a decoder $D : (\mathbb{R}^p)^T \to \mathbb{R}^n$ **corrects q errors after T steps** if it is resilient against any attack of q sensors, i.e., if for any initial condition $x(0) \in \mathbb{R}^n$, and for any attack vectors $e(0), \ldots, e(T-1)$ with support K, $|K| = q$, we have:

$$D(y(0), \ldots, y(T-1)) = x(0).$$

- We say that q **errors are correctable**, for the system (A, C), if there exists a decoder that can correct q errors.
- Note: correcting $q = 0$ errors is equivalent to observability.

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps for the pair (A, C) iff:

$$\forall x \neq 0, \ |supp(Cx) \cup supp(CAx) \cup \cdots \cup supp(CA^{T-1}x)| > 2q. \tag{1}$$

Correction of q errors

Necessary and sufficient conditions

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps for the pair (A, C) iff:

$$\forall x \neq 0, \ |supp(Cx) \cup supp(CAx) \cup \cdots \cup supp(CA^{T-1}x)| > 2q. \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to spread the components of the state x.

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps for the pair (A, C) iff:

$$\forall x \neq 0, \ |supp(Cx) \cup supp(CAx) \cup \cdots \cup supp(CA^{T-1}x)| > 2q. \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to spread the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps for the pair (A, C) iff:

$$\forall x \neq 0, \ |supp(Cx) \cup supp(CAx) \cup \cdots \cup supp(CA^{T-1}x)| > 2q. \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to spread the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

$$\text{For } x = \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} \text{ we have } Ax = \begin{bmatrix} 0 \\ x_3 \\ 0 \end{bmatrix}, \quad A^2x = \begin{bmatrix} x_3 \\ 0 \\ 0 \end{bmatrix}$$

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions

Proposition

Let $T > 0$ be fixed. Then q errors are correctable after T steps for the pair (A, C) iff:

$$\forall x \neq 0,\ |supp(Cx) \cup supp(CAx) \cup \cdots \cup supp(CA^{T-1}x)| > 2q. \tag{1}$$

- Interpretation of condition (1): $C, CA, \ldots, CA^{T-1}$ have to spread the components of the state x.
  - Example of a good pair (A, C):

$$A = \begin{bmatrix} 010 \\ 001 \\ 100 \end{bmatrix} \text{ (circular permutation)}, \quad C = \text{identity}$$

$$\text{For } x = \begin{bmatrix} 0 \\ 0 \\ x_3 \end{bmatrix} \text{ we have } Ax = \begin{bmatrix} 0 \\ x_3 \\ 0 \end{bmatrix}, \quad A^2x = \begin{bmatrix} x_3 \\ 0 \\ 0 \end{bmatrix}$$

  - Example of a very bad pair (A, C): $A =$ identity, $C =$ identity

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions (some observations)

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem);

Correction of q errors

Necessary and sufficient conditions (some observations)

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem);
- No more than $p/2$ errors can be corrected since $2q$ is necessarily smaller than p.

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions (some observations)

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem);
- No more than $p/2$ errors can be corrected since $2q$ is necessarily smaller than p.
- This is a fundamental limitation: if an attacker has access to more than half of the sensors ($> p/2$), it is impossible to reconstruct the state.

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions (some observations)

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem);
- No more than $p/2$ errors can be corrected since $2q$ is necessarily smaller than p.
- This is a fundamental limitation: if an attacker has access to more than half of the sensors ($> p/2$), it is impossible to reconstruct the state.
- But can we reconstruct the state if an attacker has access to less than half of the sensors ($< p/2$)?

CyPhyLab UCLA

Correction of q errors

Necessary and sufficient conditions (some observations)

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Number of correctable errors does not increase beyond $T = n$ steps (Cayley-Hamilton theorem);
- No more than $p/2$ errors can be corrected since $2q$ is necessarily smaller than p.
- This is a fundamental limitation: if an attacker has access to more than half of the sensors ($> p/2$), it is impossible to reconstruct the state.
- But can we reconstruct the state if an attacker has access to less than half of the sensors ($< p/2$)?

Proposition

For almost all pairs (A, C), the number of correctable errors is maximal and equal to $\lceil p/2 - 1 \rceil$.

Correction of q errors

Some observations

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Given a specific pair (A, C), what is the number q of correctable errors?

Correction of q errors

Some observations

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Given a specific pair (A, C), what is the number q of correctable errors?
- A pair (A, C) is said to be **q-sparse observable** if all the pairs (A, C'), obtained from (A, C) by removing q rows from C, remain observable.

CyPhyLab UCLA

Correction of q errors

Some observations

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Given a specific pair (A, C), what is the number q of correctable errors?
- A pair (A, C) is said to be **q-sparse observable** if all the pairs (A, C'), obtained from (A, C) by removing q rows from C, remain observable.

Theorem

For any pair (A, C), q errors are correctable iff (A, C) is $2q$-sparse observable.

CyPhyLab UCLA

Correction of q errors

Some observations

$$\forall x \neq 0,\ |\text{supp}(Cx) \cup \text{supp}(CAx) \cup \cdots \cup \text{supp}(CA^{T-1}x)| > 2q$$

- Given a specific pair (A, C), what is the number q of correctable errors?
- A pair (A, C) is said to be **q-sparse observable** if all the pairs (A, C'), obtained from (A, C) by removing q rows from C, remain observable.

Theorem

For any pair (A, C), q errors are correctable iff (A, C) is $2q$-sparse observable.

- There is a particular case that can be efficiently checked:

Proposition

Let A be a diagonalizable matrix with eigenvalues of different magnitudes. Then, for any C of compatible dimensions, q errors are correctable for the pair (A, C) iff $|supp(Cv)| > 2q$ for every eigenvector v of A.

Improving resilience by control

- What can we do if the number of correctable errors for a pair (A, C) is below $p/2$?

Improving resilience by control

- What can we do if the number of correctable errors for a pair (A, C) is below $p/2$?

Proposition

Assume the pair (A, B) to be controllable. Then, for almost any choice of n numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ with distinct magnitudes, there exists a feedback control law $u = Kx$ such that:

1. *the eigenvalues of $A + BK$ are $\lambda_1, \ldots, \lambda_n$;*
2. *the number of correctable errors after n steps for the pair $(A + BK, C)$ is maximal and equal to $\lceil p/2 - 1 \rceil$.*

Improving resilience by control

- What can we do if the number of correctable errors for a pair (A, C) is below $p/2$?

Proposition

Assume the pair (A, B) to be controllable. Then, for almost any choice of n numbers $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ with distinct magnitudes, there exists a feedback control law $u = Kx$ such that:

1. *the eigenvalues of $A + BK$ are $\lambda_1, \ldots, \lambda_n$;*
2. *the number of correctable errors after n steps for the pair $(A + BK, C)$ is maximal and equal to $\lceil p/2 - 1 \rceil$.*

This result makes sense in a scenario where the feedback control law $u = Kx$ is local and not subject to attacks.

CyPhyLab UCLA

Lecture plan

- Control under sensor attacks
- Attacking sensors
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- **State reconstruction under sensor attacks:**
  - **Convex relaxation of an ℓ_0 optimization problem**
  - Observer-based state reconstruction
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:
  - If we know $x(0)$ we know $x(t) = A^t x(0)$ and thus know $e(t) = y(t) - Cx(t)$;

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:
  - If we know $x(0)$ we know $x(t) = A^t x(0)$ and thus know $e(t) = y(t) - Cx(t)$;
  - If we know $e(t)$ then we known its support K;

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:
  - If we know $x(0)$ we know $x(t) = A^t x(0)$ and thus know $e(t) = y(t) - Cx(t)$;
  - If we know $e(t)$ then we known its support K;
  - If we know the support K, then we can use the outputs not supported on K to correctly estimate $x(0)$.

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:
  - If we know $x(0)$ we know $x(t) = A^t x(0)$ and thus know $e(t) = y(t) - Cx(t)$;
  - If we know $e(t)$ then we known its support K;
  - If we know the support K, then we can use the outputs not supported on K to correctly estimate $x(0)$.
- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{2}$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- We have received observations $y(0), \ldots, y(T-1)$ and we want to find the true initial state $x(0)$;
- Estimating $x(0)$, or $e(t)$, or K are equivalent problems:
  - If we know $x(0)$ we know $x(t) = A^t x(0)$ and thus know $e(t) = y(t) - Cx(t)$;
  - If we know $e(t)$ then we known its support K;
  - If we know the support K, then we can use the outputs not supported on K to correctly estimate $x(0)$.
- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \operatorname{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{2}$$

- In other words, among all possible attack sets $\hat{K}$ look for smallest one that is dynamically consistent with the received observations $y(0), \ldots, y(T-1)$.

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{3}$$

State reconstruction under sensor attacks

Convex relaxation approach

- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{3}$$

- How good is this decoder?

State reconstruction under sensor attacks

Convex relaxation approach

- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{3}$$

- How good is this decoder?

Proposition

If q errors are correctable for a pair (A, C), then they can be corrected by the above decoder.

State reconstruction under sensor attacks

Convex relaxation approach

- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{3}$$

- How good is this decoder?

Proposition

If q errors are correctable for a pair (A, C), then they can be corrected by the above decoder.

- This decoder is optimal in the sense that there is no decoder that is strictly better;

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Our decoder: search for the $\hat{x}(0)$ that provides the smallest $\hat{K}$, i.e., solve:

$$\begin{aligned} \underset{\hat{x}(0),\hat{K}}{\text{minimize}} \quad & |\hat{K}| \\ \text{subject to} \quad & \text{supp}(y(t) - CA^t\hat{x}(0)) \subseteq \hat{K} \quad , t = 0, \ldots, T-1. \end{aligned} \tag{3}$$

- How good is this decoder?

Proposition

If q errors are correctable for a pair (A, C), then they can be corrected by the above decoder.

- This decoder is optimal in the sense that there is no decoder that is strictly better;
- Unfortunately, problem (3) is NP-hard.

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Idea: relax the previous decoder to make it computationally tractable;
  - Inspiration from the ℓ_1-relaxation techniques used in compressed sensing and error correction over the reals.

State reconstruction under sensor attacks

Convex relaxation approach

- Idea: relax the previous decoder to make it computationally tractable;
  - Inspiration from the ℓ_1-relaxation techniques used in compressed sensing and error correction over the reals.
- Some notation first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\begin{bmatrix} y(0) & \dots & y(T-1) \end{bmatrix}}_{Y \in \mathbb{R}^{p \times T}} = \underbrace{\begin{bmatrix} Cx & \dots & CA^{T-1}x \end{bmatrix}}_{\Phi x} + \underbrace{\begin{bmatrix} e(0) & \dots & e(T-1) \end{bmatrix}}_{E \in \mathbb{R}^{p \times T}}$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Idea: relax the previous decoder to make it computationally tractable;
  - Inspiration from the ℓ_1-relaxation techniques used in compressed sensing and error correction over the reals.
- Some notation first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\begin{bmatrix} y(0) & \dots & y(T-1) \end{bmatrix}}_{Y \in \mathbb{R}^{p \times T}} = \underbrace{\begin{bmatrix} Cx & \dots & CA^{T-1}x \end{bmatrix}}_{\Phi x} + \underbrace{\begin{bmatrix} e(0) & \dots & e(T-1) \end{bmatrix}}_{E \in \mathbb{R}^{p \times T}}$$

- Define the ℓ_0 norm of E as the number of nonzero rows of E (number of attacked sensors):

$$\|E\|_{\ell_0} = |\text{rowsupp}(E)|$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Idea: relax the previous decoder to make it computationally tractable;
  - Inspiration from the ℓ_1-relaxation techniques used in compressed sensing and error correction over the reals.
- Some notation first:
  - Collect observations from $t = 0$ to $t = T - 1$ in a $p \times T$ matrix:

$$\underbrace{\begin{bmatrix} y(0) & \dots & y(T-1) \end{bmatrix}}_{Y \in \mathbb{R}^{p \times T}} = \underbrace{\begin{bmatrix} Cx & \dots & CA^{T-1}x \end{bmatrix}}_{\Phi x} + \underbrace{\begin{bmatrix} e(0) & \dots & e(T-1) \end{bmatrix}}_{E \in \mathbb{R}^{p \times T}}$$

- Define the ℓ_0 norm of E as the number of nonzero rows of E (number of attacked sensors):

$$\|E\|_{\ell_0} = |\text{rowsupp}\,(E)|$$

- Let's rewrite the previous "*unbeatable*" decoder using this notation:
  - smallest number of attacked sensors that explain the received observations:

$$\underset{x}{\text{minimize}} \; \| \underbrace{Y - \Phi x}_{E} \|_{\ell_0}$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- ℓ_0 decoder (NP-hard):

$$\underset{x}{\text{minimize}} \left\| Y - \Phi x \right\|_{\ell_0}$$

State reconstruction under sensor attacks

Convex relaxation approach

- ℓ_0 decoder (NP-hard):

$$\underset{x}{\text{minimize}} \left\| Y - \Phi x \right\|_{\ell_0}$$

- Relaxation idea: instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable):

State reconstruction under sensor attacks

Convex relaxation approach

- ℓ_0 decoder (NP-hard):

$$\underset{x}{\text{minimize}} \; \|Y - \Phi x\|_{\ell_0}$$

- Relaxation idea: instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable):
  - replace number of nonzero rows of E by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of E:

$$\underset{x}{\text{minimize}} \; \| \underbrace{Y - \Phi x}_{\in \mathbb{R}^{p \times T}} \|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \| \underbrace{(Y - \Phi x)_i}_{\in \mathbb{R}^{T}} \|_{\ell_r}$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- ℓ_0 decoder (NP-hard):

$$\underset{x}{\text{minimize}} \; \|Y - \Phi x\|_{\ell_0}$$

- Relaxation idea: instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable):
  - replace number of nonzero rows of E by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of E:

$$\underset{x}{\text{minimize}} \; \|\underbrace{Y - \Phi x}_{\in \mathbb{R}^{p \times T}}\|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \|\underbrace{(Y - \Phi x)_i}_{\in \mathbb{R}^T}\|_{\ell_r}$$

- **Magnitude** of a row of E measured by its ℓ_r norm (in $\mathbb{R}^T$), for any $r \geq 1$.

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- ℓ_0 decoder (NP-hard):

$$\underset{x}{\text{minimize}} \left\| Y - \Phi x \right\|_{\ell_0}$$

- Relaxation idea: instead of "ℓ_0 norm" (intractable), use ℓ_1 norm (convex program, tractable):
  - replace number of nonzero rows of E by $\underbrace{\text{sum}}_{\ell_1}$ of the $\underbrace{\textbf{magnitudes}}_{\ell_r}$ of the rows of E:

$$\underset{x}{\text{minimize}} \| \underbrace{Y - \Phi x}_{\in \mathbb{R}^{p \times T}} \|_{\ell_1/\ell_r} = \sum_{i=1}^{p} \| \underbrace{(Y - \Phi x)_i}_{\in \mathbb{R}^T} \|_{\ell_r}$$

- **Magnitude** of a row of E measured by its ℓ_r norm (in $\mathbb{R}^T$), for any $r \geq 1$.

- $\ell_0 \to \ell_1$ relaxation idea used in compressed sensing (recovery of sparse signals), and error correction over the reals (cf. Candes, Tao, Donoho, etc.).

CyPhyLab UCLA

State reconstruction under sensor attacks

Convex relaxation approach

- Randomly generated system (A, C) with $n = 30$ and $p = 20$ (Gaussian entries)
- Used ℓ_1/ℓ_2 decoder

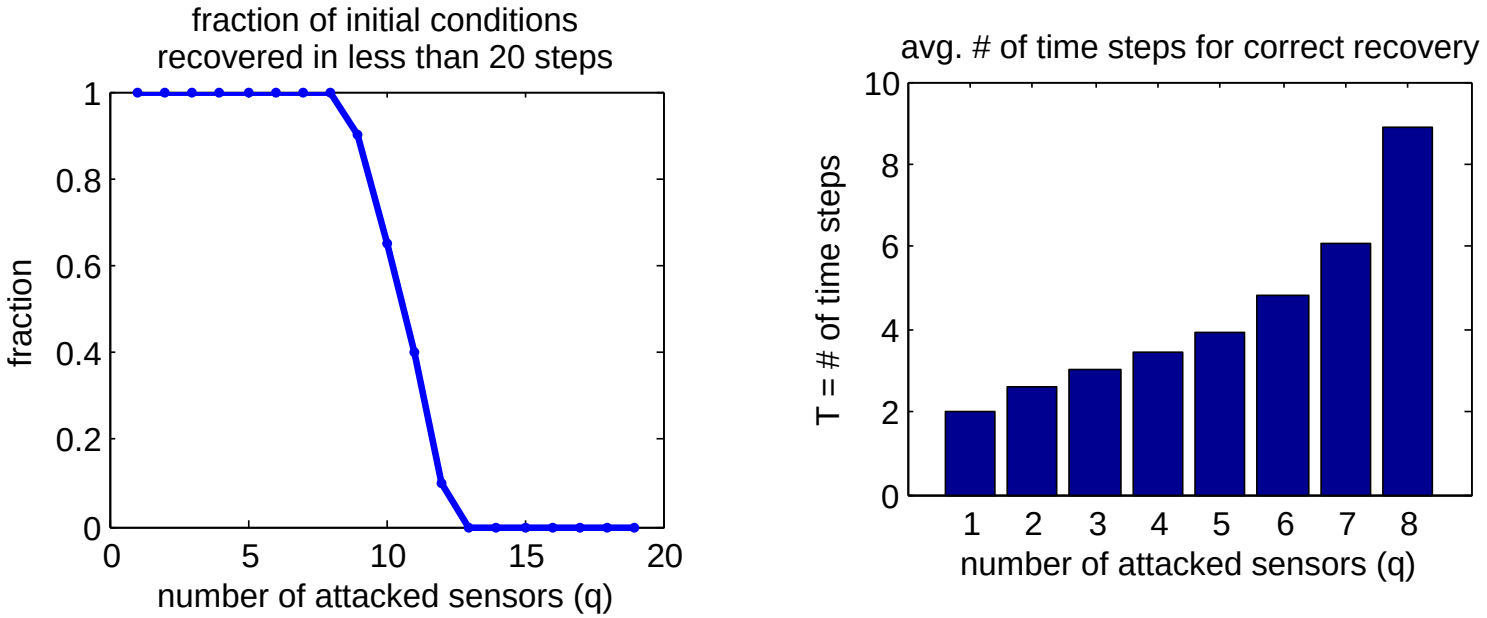


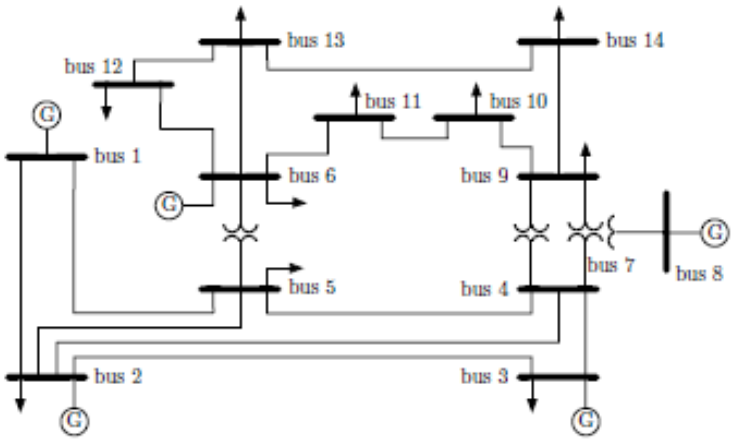
Figure: (a) Fraction of initial conditions (out of 20) that were correctly recovered in less than $T = 20$ time steps, for different values of q. (b) Average number of time steps it took to correctly recover the initial state as a function of the number of corrupted components.

State reconstruction under sensor attacks

Convex relaxation approach

Electric power network: IEEE 14-bus power network (5 generators, 14 buses)

- $n = 2 \times 5 = 10$ states for the rotor angles δ_i and the frequencies $d\delta_i/dt$ of each generator i
- $p = 35$ sensors to measure: real power injections at every bus (14 sensors), real power flows along every branch (20 sensors), rotor angle at generator 1 (1 sensor) [1]



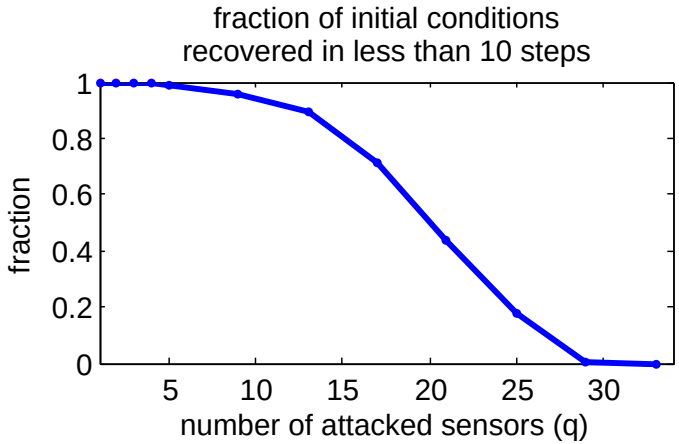


Figure: (a) IEEE 14-bus power network (b) Fraction of initial conditions that were correctly recovered in less than $T = 10$ steps. For each value of q, 200 simulations were carried out with different initial conditions and different sets of attacked sensors using the ℓ_1/ℓ_∞ decoder.

Lecture plan

- Control under sensor attacks
- Attacking sensors
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- **State reconstruction under sensor attacks:**
  - Convex relaxation of an ℓ_0 optimization problem
  - **Observer-based state reconstruction**
  - Satisfiability Modulo Theories approach

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

The ℓ_1/ℓ_r decoder needs to:

1. Collect T measurements;
2. Solve an optimization problem to obtain $\hat{x}(0)$;
3. Use the dynamics $x(t+1) = Ax(t)$ to obtain the estimate of the state $\hat{x}(T-1) = A^{T-1}\hat{x}(0)$ at the current time $T-1$.

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

The ℓ_1/ℓ_r decoder needs to:

1. Collect T measurements;
2. Solve an optimization problem to obtain $\hat{x}(0)$;
3. Use the dynamics $x(t+1) = Ax(t)$ to obtain the estimate of the state $\hat{x}(T-1) = A^{T-1}\hat{x}(0)$ at the current time $T-1$.

This process suffers from:

1. Being slow (if using an off-the-shelf optimization engine);
2. Poor real-time guarantees;
3. Not secure!

State reconstruction under sensor attacks

Observer-based approach

The ℓ_1/ℓ_r decoder needs to:

1. Collect T measurements;
2. Solve an optimization problem to obtain $\hat{x}(0)$;
3. Use the dynamics $x(t+1) = Ax(t)$ to obtain the estimate of the state $\hat{x}(T-1) = A^{T-1}\hat{x}(0)$ at the current time $T-1$.

This process suffers from:

1. Being slow (if using an off-the-shelf optimization engine);
2. Poor real-time guarantees;
3. Not secure!

When security is not a concern, one typically uses an observer (Luenberger) or an estimator (Kalman) to reconstruct/estimate the state.

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

The ℓ_1/ℓ_r decoder needs to:

1. Collect T measurements;
2. Solve an optimization problem to obtain $\hat{x}(0)$;
3. Use the dynamics $x(t+1) = Ax(t)$ to obtain the estimate of the state $\hat{x}(T-1) = A^{T-1}\hat{x}(0)$ at the current time $T-1$.

This process suffers from:

1. Being slow (if using an off-the-shelf optimization engine);
2. Poor real-time guarantees;
3. Not secure!

When security is not a concern, one typically uses an observer (Luenberger) or an estimator (Kalman) to reconstruct/estimate the state.

Question: can we devise an observer that is resilient to attacks?

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

- Let us extend the original dynamics to also model the evolution of sequences of attacks and observations of length T:

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t-1) &= Qz(t-1) \end{aligned}$$

$$z(t) = \begin{bmatrix} x(t-T+1) \\ a(t-T+1) \\ \vdots \\ a(t-1) \\ a(t) \end{bmatrix}, \quad Y(t-1) = \begin{bmatrix} y(t-T) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix}.$$

State reconstruction under sensor attacks

Observer-based approach

- Let us extend the original dynamics to also model the evolution of sequences of attacks and observations of length T:

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t-1) &= Qz(t-1) \end{aligned}$$

$$z(t) = \begin{bmatrix} x(t-T+1) \\ a(t-T+1) \\ \vdots \\ a(t-1) \\ a(t) \end{bmatrix}, \quad Y(t-1) = \begin{bmatrix} y(t-T) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix}.$$

- How can we model the evolution of the attack vector a if the attack is unknown?

State reconstruction under sensor attacks

Observer-based approach

- Let us extend the original dynamics to also model the evolution of sequences of attacks and observations of length T:

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t-1) &= Qz(t-1) \end{aligned}$$

$$z(t) = \begin{bmatrix} x(t-T+1) \\ a(t-T+1) \\ \vdots \\ a(t-1) \\ a(t) \end{bmatrix}, \quad Y(t-1) = \begin{bmatrix} y(t-T) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix}.$$

- How can we model the evolution of the attack vector a if the attack is unknown?

$$a(t) = y(t) - Cx(t)$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

- Let us extend the original dynamics to also model the evolution of sequences of attacks and observations of length T:

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t-1) &= Qz(t-1) \end{aligned}$$

$$z(t) = \begin{bmatrix} x(t-T+1) \\ a(t-T+1) \\ \vdots \\ a(t-1) \\ a(t) \end{bmatrix}, \quad Y(t-1) = \begin{bmatrix} y(t-T) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix}.$$

- How can we model the evolution of the attack vector a if the attack is unknown?

$$a(t) = y(t) - Cx(t) = y(t) - C(Ax(t-1) + Bu(t-1))$$

State reconstruction under sensor attacks

Observer-based approach

- Let us extend the original dynamics to also model the evolution of sequences of attacks and observations of length T:

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t-1) &= Qz(t-1) \end{aligned}$$

$$z(t) = \begin{bmatrix} x(t-T+1) \\ a(t-T+1) \\ \vdots \\ a(t-1) \\ a(t) \end{bmatrix}, \quad Y(t-1) = \begin{bmatrix} y(t-T) \\ \vdots \\ y(t-2) \\ y(t-1) \end{bmatrix}.$$

- How can we model the evolution of the attack vector a if the attack is unknown?

$$a(t) = y(t) - Cx(t) = y(t) - C(Ax(t-1) + Bu(t-1))$$

By making $y(t)$ part of the input $u'(t-1)$, we can express $a(t)$ as a linear combination of $z(t-1)$ and $u'(t-1)$.

State reconstruction under sensor attacks

Observer-based approach

- A Luenberger observer is now given by:

$$\begin{aligned} \hat{z}(t) &= A'\hat{z}(t-1) + B'u'(t-1) + L(Y(t-1) - \hat{Y}(t-1)) \\ \hat{Y}(t-1) &= Q\hat{z}(t-1) \end{aligned}$$

State reconstruction under sensor attacks

Observer-based approach

- A Luenberger observer is now given by:

$$\begin{aligned} \hat{z}(t) &= A'\hat{z}(t-1) + B'u'(t-1) + L(Y(t-1) - \hat{Y}(t-1)) \\ \hat{Y}(t-1) &= Q\hat{z}(t-1) \end{aligned}$$

- It is convenient to see the dynamics of the Luenberger observer as the alternation of two steps:

$$\begin{aligned} \overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\ \hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \end{aligned}$$

State reconstruction under sensor attacks

Observer-based approach

- A Luenberger observer is now given by:

$$\begin{aligned} \hat{z}(t) &= A'\hat{z}(t-1) + B'u'(t-1) + L(Y(t-1) - \hat{Y}(t-1)) \\ \hat{Y}(t-1) &= Q\hat{z}(t-1) \end{aligned}$$

- It is convenient to see the dynamics of the Luenberger observer as the alternation of two steps:

$$\begin{aligned} \overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\ \hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \end{aligned}$$

- In general, $\hat{a}(t-1), \ldots, \hat{a}(t-T)$ are not supported on an attack set K with $|K| = q$.

State reconstruction under sensor attacks

Observer-based approach

- This can be resolved by projecting $\hat{z}$ on the set of vectors containing $p - q$ zero elements.

$$
\begin{aligned}
\overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\
\hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \\
\hat{z}(t) &= \Pi(\hat{z}(t)) && \text{(projection on constraint set)}
\end{aligned}
$$

State reconstruction under sensor attacks

Observer-based approach

- This can be resolved by projecting $\hat{z}$ on the set of vectors containing $p - q$ zero elements.

$$
\begin{aligned}
\overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\
\hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \\
\hat{z}(t) &= \Pi(\hat{z}(t)) && \text{(projection on constraint set)}
\end{aligned}
$$

- If we interpret the measurement update step as a gradient step, the above equations describe a recursive version of the projected gradient algorithm (hard thresholding, proximal gradient, ...).

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

- This can be resolved by projecting $\hat{z}$ on the set of vectors containing $p - q$ zero elements.

$$\begin{aligned} \overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\ \hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \\ \hat{z}(t) &= \Pi(\hat{z}(t)) && \text{(projection on constraint set)} \end{aligned}$$

- If we interpret the measurement update step as a gradient step, the above equations describe a recursive version of the projected gradient algorithm (hard thresholding, proximal gradient, ...).
- Unfortunately, this observer only converges under restrictive assumptions* on the measurement equation $Y = Qz$.

* *Iterative hard thresholding for compressed sensing.*
T. Blumensath, M. E. Davies.
Applied and Computational Harmonic Analysis, 27(3), 2009.

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

- This can be resolved by projecting $\hat{z}$ on the set of vectors containing $p - q$ zero elements.

$$\begin{aligned} \overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\ \hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \\ \hat{z}(t) &= \Pi(\hat{z}(t)) && \text{(projection on constraint set)} \end{aligned}$$

- If we interpret the measurement update step as a gradient step, the above equations describe a recursive version of the projected gradient algorithm (hard thresholding, proximal gradient, ...).
- Unfortunately, this observer only converges under restrictive assumptions* on the measurement equation $Y = Qz$.
- Solution: repeat the measurement update step several times to compensate for the "damage" caused by the time update and projection steps.

* *Iterative hard thresholding for compressed sensing.*
T. Blumensath, M. E. Davies.
Applied and Computational Harmonic Analysis, 27(3), 2009.

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

- This can be resolved by projecting $\hat{z}$ on the set of vectors containing $p - q$ zero elements.

$$\begin{aligned} \overline{z}(t-1) &= \hat{z}(t-1) + L'(Y(t-1) - \hat{Y}(t-1)) && \text{(measurement update)} \\ \hat{z}(t) &= A'\overline{z}(t-1) + B'u'(t-1) && \text{(time update)} \\ \hat{z}(t) &= \Pi(\hat{z}(t)) && \text{(projection on constraint set)} \end{aligned}$$

- If we interpret the measurement update step as a gradient step, the above equations describe a recursive version of the projected gradient algorithm (hard thresholding, proximal gradient, ...).
- Unfortunately, this observer only converges under restrictive assumptions* on the measurement equation $Y = Qz$.
- Solution: repeat the measurement update step several times to compensate for the "damage" caused by the time update and projection steps.
- We call this observer the **Event-Triggered Projected Luenberger (ETPL)** observer.

** Iterative hard thresholding for compressed sensing.*
T. Blumensath, M. E. Davies.
Applied and Computational Harmonic Analysis, 27(3), 2009.

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

$$\begin{aligned} z(t) &= A' z(t-1) + B' u'(t-1) \\ Y(t) &= Qz(t-1) \end{aligned}$$

- Let δ_{2q} be the smallest eigenvalue of all the matrices ${Q'}^T Q'$ with Q' obtained from Q by removing $2q$ sensors.

State reconstruction under sensor attacks

Observer-based approach

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t) &= Qz(t-1) \end{aligned}$$

- Let δ_{2q} be the smallest eigenvalue of all the matrices $Q'^T Q'$ with Q' obtained from Q by removing $2q$ sensors.

Theorem

For any $2q$-sparse observable linear control system with $\delta_{2q} > \frac{4}{9}\lambda_{\max}(Q^T Q)$, the ETPL observer converges to the true state and attack sequence whenever:

1. *$L = Q^T \Sigma$ for any positive definite matrix Σ satisfying $\lambda_{max}(\Sigma) < \lambda^{-1}_{max}(Q^T Q)$;*
2. *the attack vectors are supported on a set K with $|K| \leq q$.*

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

$$\begin{aligned} z(t) &= A'z(t-1) + B'u'(t-1) \\ Y(t) &= Qz(t-1) \end{aligned}$$

- Let δ_{2q} be the smallest eigenvalue of all the matrices $Q'^T Q'$ with Q' obtained from Q by removing $2q$ sensors.

Theorem

For any $2q$-sparse observable linear control system with $\delta_{2q} > \frac{4}{9}\lambda_{\max}(Q^T Q)$, the ETPL observer converges to the true state and attack sequence whenever:

1. *$L = Q^T \Sigma$ for any positive definite matrix Σ satisfying $\lambda_{max}(\Sigma) < \lambda^{-1}_{max}(Q^T Q)$;*
2. *the attack vectors are supported on a set K with $|K| \leq q$.*

- Does it actually work?

CyPhyLab UCLA

State reconstruction under sensor attacks

Observer-based approach

Lecture plan

- Control under sensor attacks
- Attacking sensors
- Some fundamental results
  - Separation between estimation and control under sensor attacks
  - Possibility/impossibility results
- **State reconstruction under sensor attacks:**
  - Convex relaxation of an ℓ_0 optimization problem
  - Observer-based state reconstruction
  - **Satisfiability Modulo Theories approach**

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

- The relaxation approach as well as the observer-based approach are efficient but do not always correct q errors when q errors are correctable.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

- The relaxation approach as well as the observer-based approach are efficient but do not always correct q errors when q errors are correctable.
- Can we devise a solution that always corrects the maximal number of correctable errors?

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

- The relaxation approach as well as the observer-based approach are efficient but do not always correct q errors when q errors are correctable.
- Can we devise a solution that always corrects the maximal number of correctable errors?
- While being efficient?

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

- The relaxation approach as well as the observer-based approach are efficient but do not always correct q errors when q errors are correctable.
- Can we devise a solution that always corrects the maximal number of correctable errors?
- While being efficient?
- We will use a SAT solver to handle the discrete part of the problem and convex optimization to handle the continuous part of the problem.

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) \end{cases}$$

Collect τ measurements:

$$\underbrace{\begin{bmatrix} y_i(t-\tau+1) \\ y_i(t-\tau) \\ \vdots \\ y_i(t) \end{bmatrix}}_{Y_i} = \underbrace{\begin{bmatrix} C_i \\ C_iA \\ \vdots \\ C_iA^{\tau-1} \end{bmatrix}}_{\mathcal{O}_i} x + \underbrace{\begin{bmatrix} a_i(t-\tau+1) \\ a_i(t-\tau) \\ \vdots \\ a_i(t) \end{bmatrix}}_{E_i}$$

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

Collect τ measurements:

$$Y_i = \begin{cases} \mathcal{O}_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ \mathcal{O}_i x & \text{if sensor } i \text{ is attack-free} \end{cases}$$

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

Collect τ measurements:

$$Y_i = \begin{cases} \mathcal{O}_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ \mathcal{O}_i x & \text{if sensor } i \text{ is attack-free} \end{cases}$$

- For each individual sensor, we define a binary indicator variable $b_i \in \mathbb{B}$ by declaring $b_i = 1$ when the ith sensor is under attack and $b_i = 0$ otherwise.

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

Collect τ measurements:

$$Y_i = \begin{cases} \mathcal{O}_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ \mathcal{O}_i x & \text{if sensor } i \text{ is attack-free} \end{cases}$$

- For each individual sensor, we define a binary indicator variable $b_i \in \mathbb{B}$ by declaring $b_i = 1$ when the ith sensor is under attack and $b_i = 0$ otherwise.

Problem

(Secure State Estimation) *For the linear control system under attack Σ_a, construct an estimate $\eta = (x, b) \in \mathbb{R}^n \times \mathbb{B}^p$ such that $\eta \models \phi$, i.e., η satisfies ϕ, where ϕ is defined as:*

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow Y_i = \mathcal{O}_i x \right) \quad \bigwedge \quad \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

Collect τ measurements:

$$Y_i = \begin{cases} \mathcal{O}_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ \mathcal{O}_i x & \text{if sensor } i \text{ is attack-free} \end{cases}$$

- For each individual sensor, we define a binary indicator variable $b_i \in \mathbb{B}$ by declaring $b_i = 1$ when the ith sensor is under attack and $b_i = 0$ otherwise.

Problem

(Secure State Estimation) *For the linear control system under attack Σ_a, construct an estimate $\eta = (x, b) \in \mathbb{R}^n \times \mathbb{B}^p$ such that $\eta \models \phi$, i.e., η satisfies ϕ, where ϕ is defined as:*

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right) \quad \bigwedge \quad \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$

Collect τ measurements:

$$Y_i = \begin{cases} \mathcal{O}_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ \mathcal{O}_i x & \text{if sensor } i \text{ is attack-free} \end{cases}$$

- For each individual sensor, we define a binary indicator variable $b_i \in \mathbb{B}$ by declaring $b_i = 1$ when the ith sensor is under attack and $b_i = 0$ otherwise.
- In the context of decision procedures on the reals, we resort to the notion of δ-completeness, i.e., we understand the inequality as follows $\| Y_i - \mathcal{O}_i x \|_2^2 \leq \delta$.

Problem

(Secure State Estimation) *For the linear control system under attack Σ_a, construct an estimate $\eta = (x, b) \in \mathbb{R}^n \times \mathbb{B}^p$ such that $\eta \models \phi$, i.e., η satisfies ϕ, where ϕ is defined as:*

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right) \quad \bigwedge \quad \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$

IMHOTEP-SMT Engine

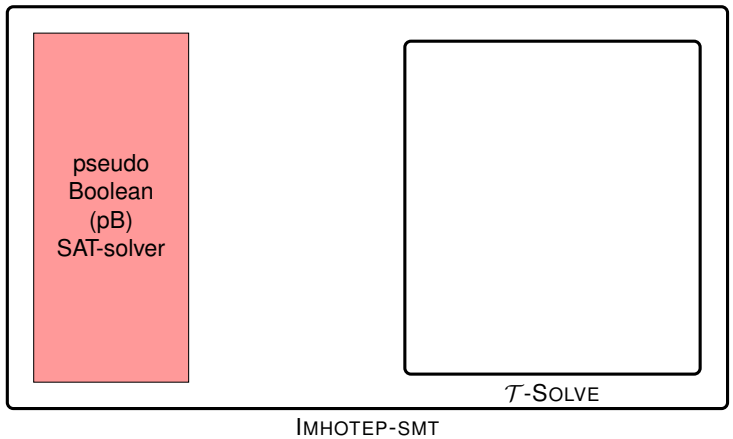
Imhotep “emmo-tep” (meaning: the one who comes in peace, is with peace) was an Egyptian mathematician, engineer, architect and physician. He was the designer of the first pyramid in Egypt.

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture I

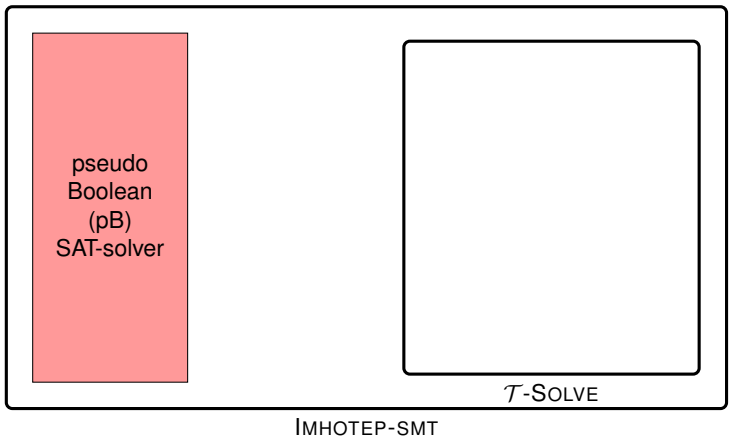
- SMT = pB-SAT solver + $\mathcal{T}$-Solver.



State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture I

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.

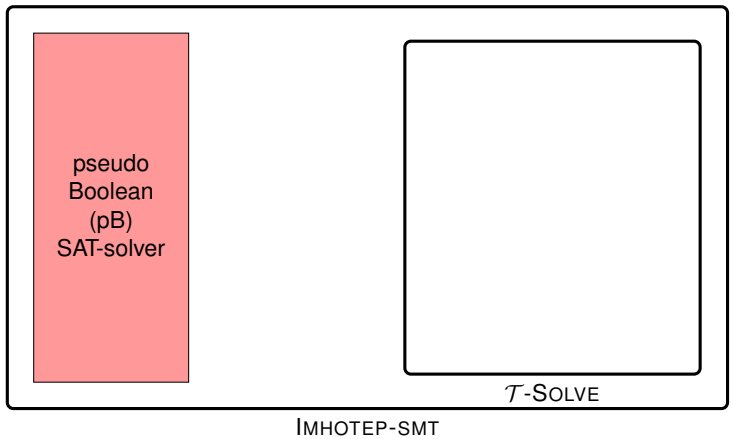


State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture I

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
  - Original formula:

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right)$$
$$\bigwedge \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$



State reconstruction under sensor attacks

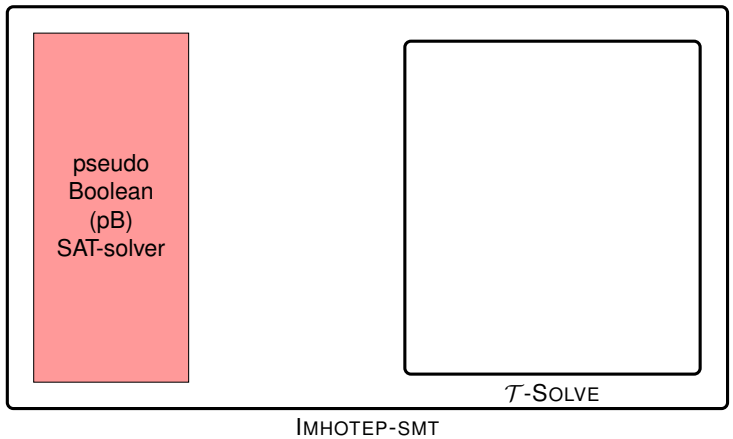
Satisfiability Modulo Theories approach: Lazy SMT Architecture I

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
  - Original formula:

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right)$$
$$\bigwedge \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$

  - Replace non-boolean variables with boolean ones

$$\phi_{initial} ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow c_i \right) \bigwedge \left(\sum_{i=1}^{p} b_i \leq s \right)$$



State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture I

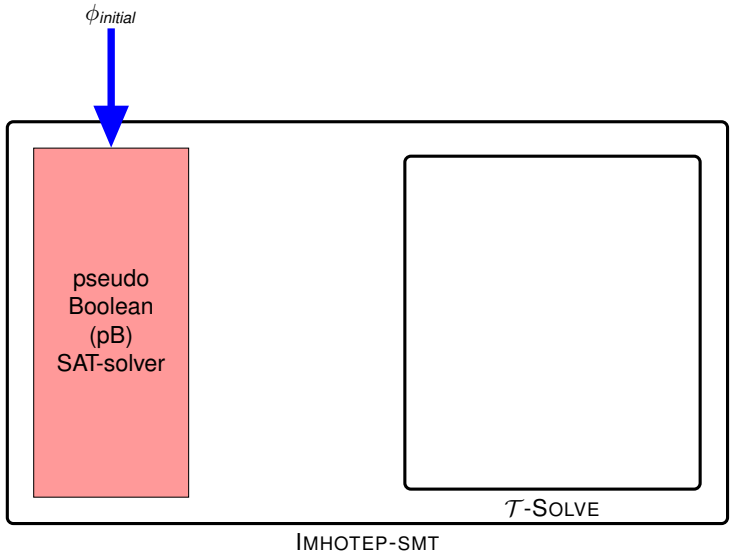
- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
  - Original formula:

$$\phi ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right)$$
$$\bigwedge \left(\sum_{i \in 1}^{p} b_i \leq s \right).$$

  - Replace non-boolean variables with boolean ones

$$\phi_{initial} ::= \bigwedge_{i=1}^{p} \left(\neg b_i \Rightarrow c_i \right) \bigwedge \left(\sum_{i=1}^{p} b_i \leq s \right)$$

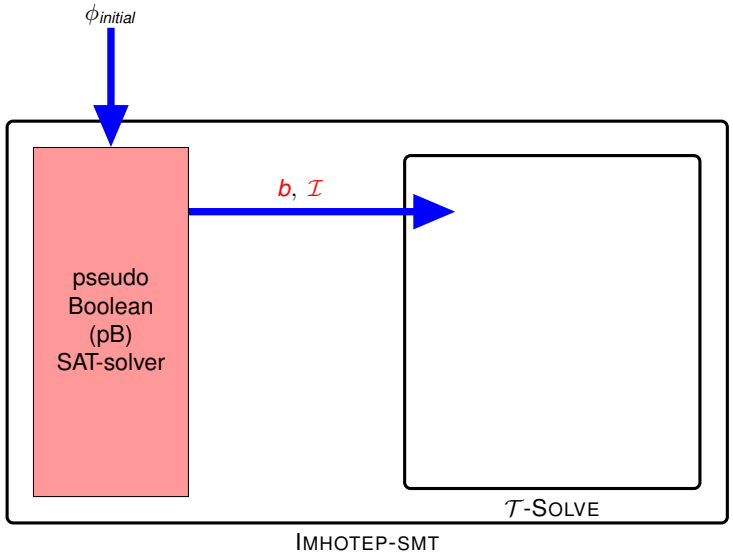
  - Pass $\phi_{initial}$ to the SAT solver.



State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture II

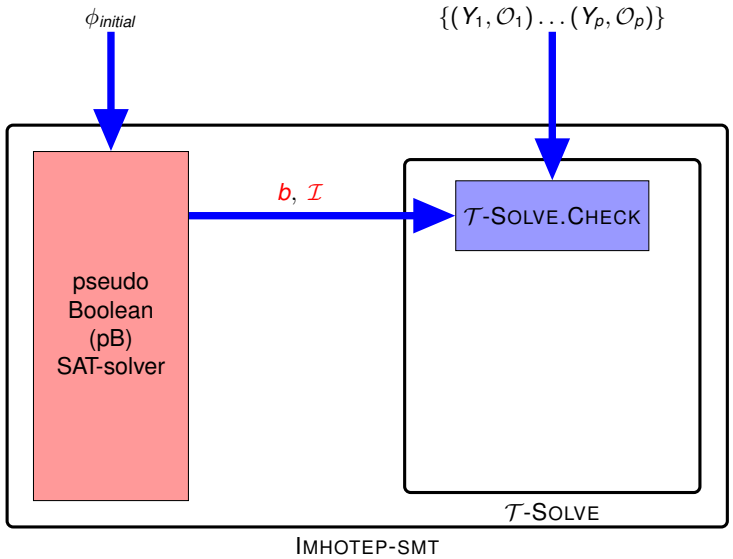
- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.



State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture II

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.

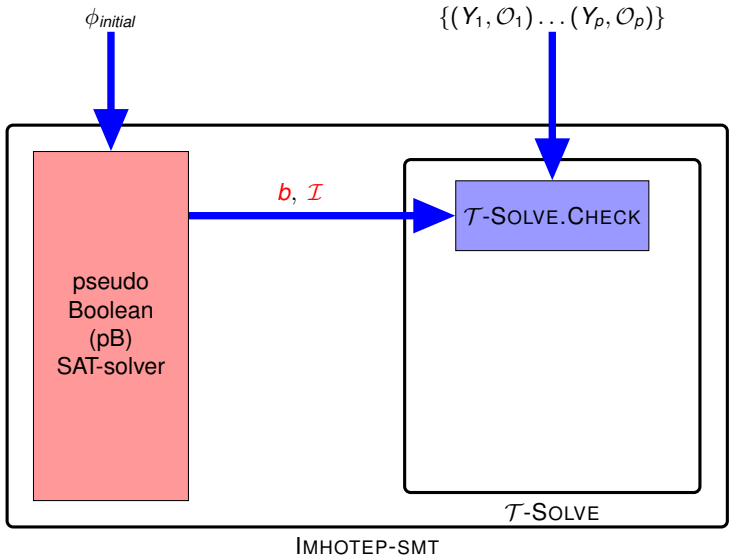


State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture II

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.

1: **Solve:**
$x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_{\mathcal{I}} - \mathcal{O}_{\mathcal{I}} x \|_2^2$



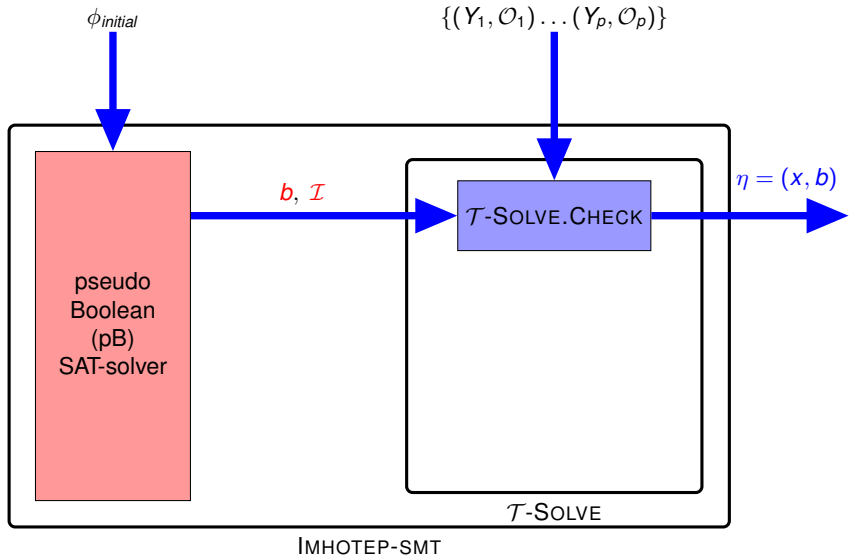
State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture II

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.

1: **Solve:**
   $x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_{\mathcal{I}} - \mathcal{O}_{\mathcal{I}} x \|_2^2$
2: **if** $\| Y_{\mathcal{I}} - \mathcal{O}_{\mathcal{I}} x \|_2^2 = 0$ **then**
3: status = SAT; ☺

6: **end if**
7: **return** (status, x);

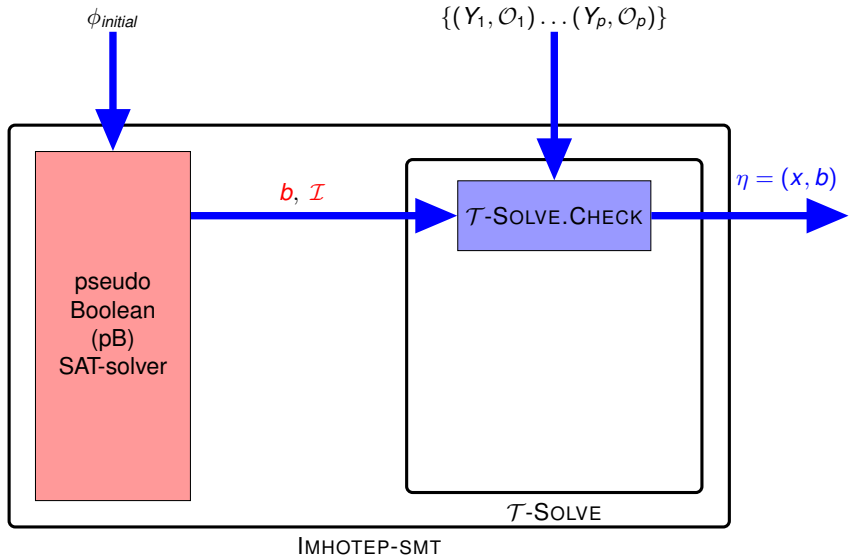


State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture II

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.

1: **Solve:**
   $x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_{\mathcal{I}} - \mathcal{O}_{\mathcal{I}} x \|_2^2$
2: **if** $\| Y_{\mathcal{I}} - \mathcal{O}_{\mathcal{I}} x \|_2^2 = 0$ **then**
3: status = SAT; ☺
4: **else**
5: status = UNSAT; ☹
6: **end if**
7: **return** (status, x);

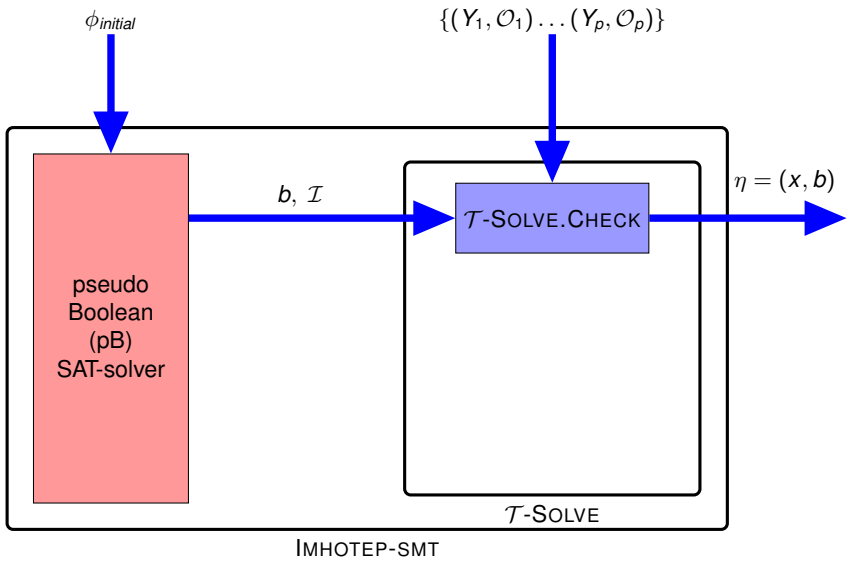


CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture III

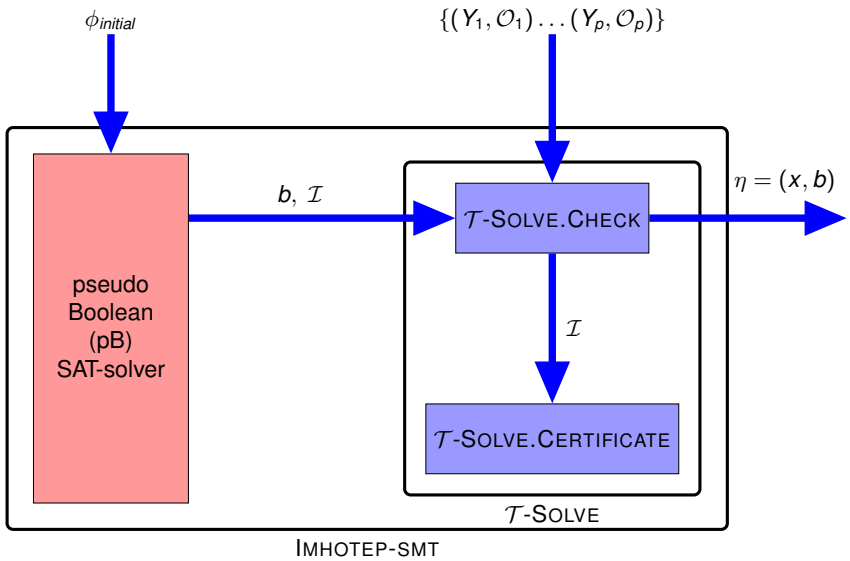
- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.



State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture III

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.
- Generate "Theory lemma" / "counter examples" / "UNSAT certificate".

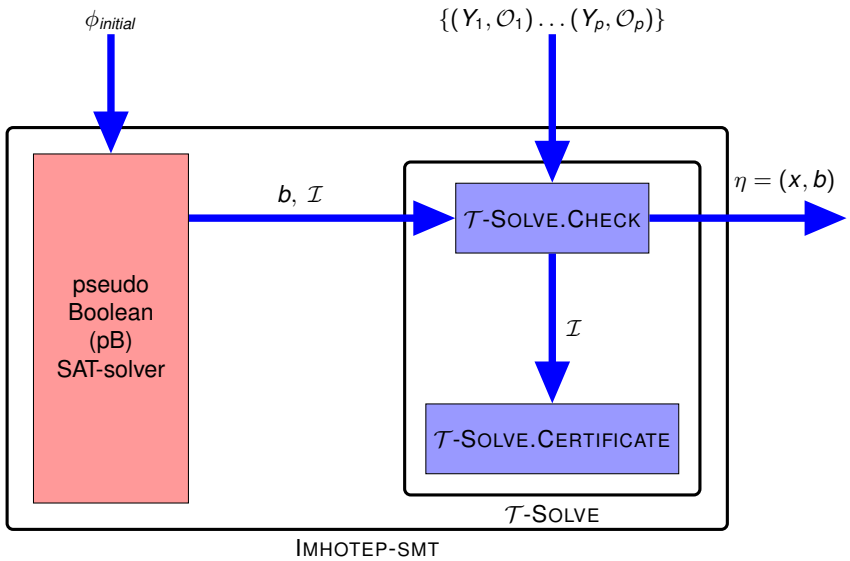


State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Lazy SMT Architecture III

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.
- Generate "Theory lemma" / "counter examples" / "UNSAT certificate".

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$



State reconstruction under sensor attacks

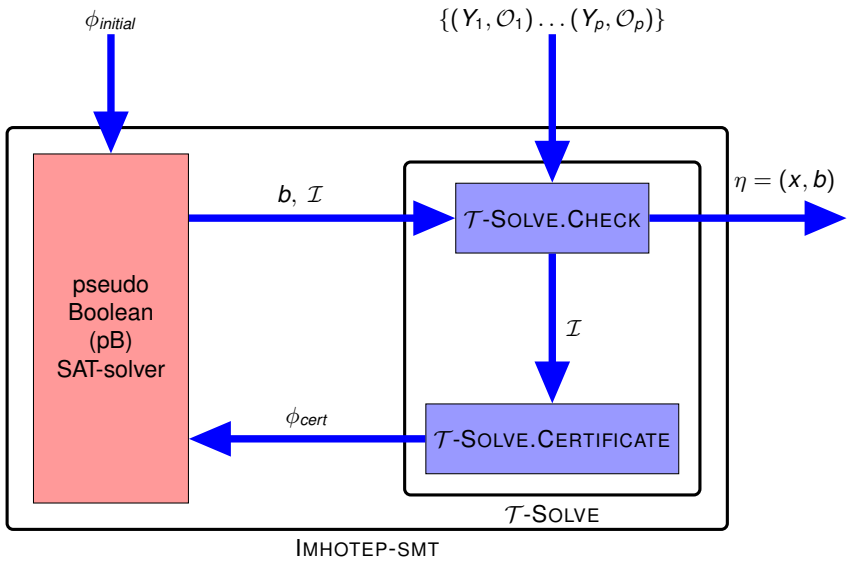
Satisfiability Modulo Theories approach: Lazy SMT Architecture III

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.
- Generate "Theory lemma" / "counter examples" / "UNSAT certificate".

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$

- Add this "certificate" to the original constraints:

$$\phi := \phi_{\text{initial}} \wedge \phi_{\text{triv-cert}}$$



CyPhyLab UCLA

State reconstruction under sensor attacks

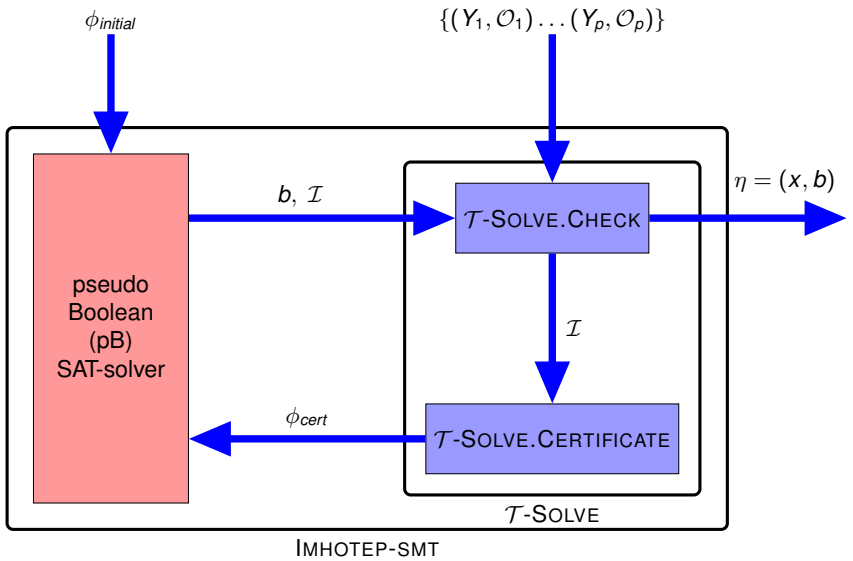
Satisfiability Modulo Theories approach: Lazy SMT Architecture III

- SMT = pB-SAT solver + $\mathcal{T}$-Solver.
- pB-SAT solver: solves the "boolean version" of the problem.
- pB-SAT solver returns an assignment for the variable b.
- We extract which sensors are "hypothesized" to be attack free $\mathcal{I}$.
- Check this assignment.
- Generate "Theory lemma" / "counter examples" / "UNSAT certificate".

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$

- Add this "certificate" to the original constraints:

$$\phi := \phi_{\text{initial}} \wedge \phi_{\text{triv-cert}}$$



REPEAT

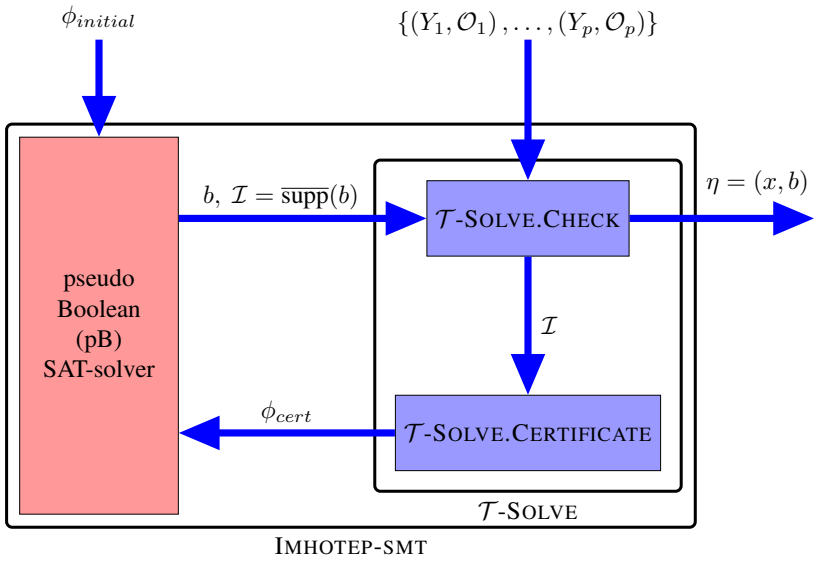
CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Termination and performance

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = Ax(t) + Bu(t), \\ y(t) & = Cx(t) + a(t) \end{cases}$$



Proposition

Let the linear dynamical system Σ_a be 2s-sparse observable. Then, IMHOTEP-SMT ***terminates*** ☺ *with:*

- $supp(b^*) \subseteq supp(b)$. ☺.
- $\|x^* - x\|_2^2 = 0$. ☺

Moreover, the number of iterations is upper bounded by $\sum_{s=0}^{s} \binom{p}{s}$. ☹

x^* is the actual system state $supp(b^*)$ is the index of the sensors under attack.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: UNSAT certificates

- Why is performance bad?

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: UNSAT certificates

- Why is performance bad?

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$

- To enhance performance, we need to generate *compact certificates*.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: UNSAT certificates

- Why is performance bad?

$$\phi_{\text{triv-cert}} = \sum_{i \in \mathcal{I}} b_i \geq 1$$

- To enhance performance, we need to generate *compact certificates*.

Lemma

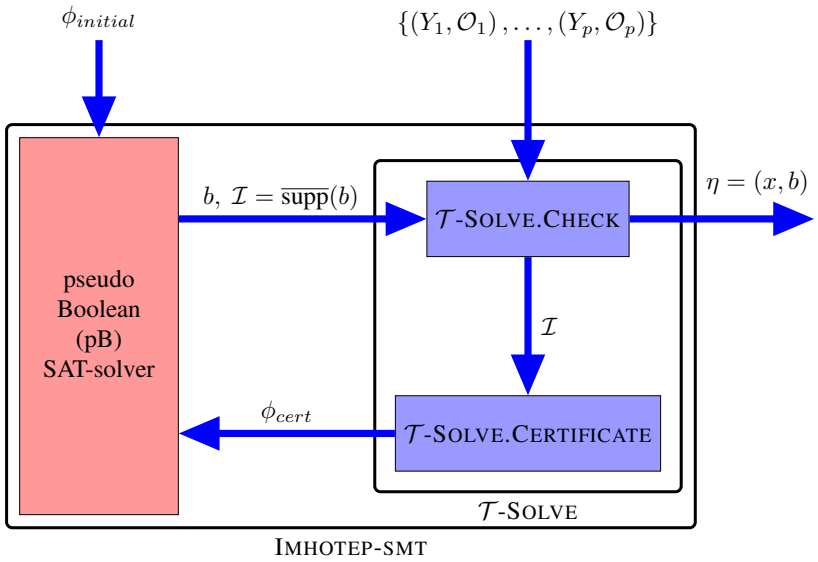
Let the linear dynamical system Σ_a be $2s$-sparse observable. If $\mathcal{T}$-SOLVE.CHECK($\mathcal{I}$) is `UNSAT` *then there exists a subset $\mathcal{I} \subset supp(b)$ with $|\mathcal{I}| \leq p - 2s + 1$ such that $\mathcal{T}$-SOLVE.CHECK($\mathcal{I}_{temp}$) is also* `UNSAT`*.*

- Trivial certificates have $p - s$ sensors.
- The proof of this lemma is constructive.
- In practice we can do better by exploiting the convex geometry.

CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: UNSAT certificates



Theorem

Let the linear dynamical system Σ_a be 2s-sparse observable. Then, IMHOTEP-SMT *terminates* ☺ *with:*

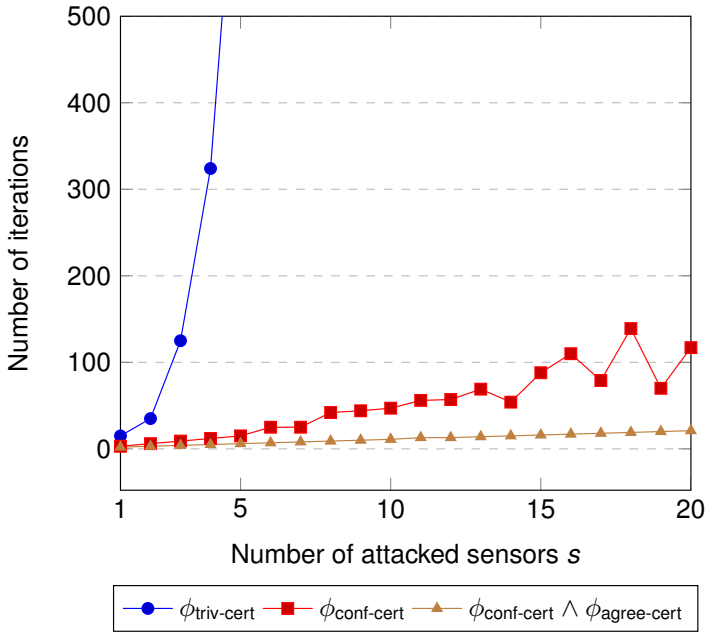
- $supp(b^*) \subseteq supp(b)$. ☺.
- $\|x^* - x\|_2 = 0$. ☺

Moreover, the number of iterations is upper bounded by $\binom{p}{p-2s+1}$. *Compare to:* $\sum_{s'=0}^{s} \binom{p}{s'}$.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Simulation results

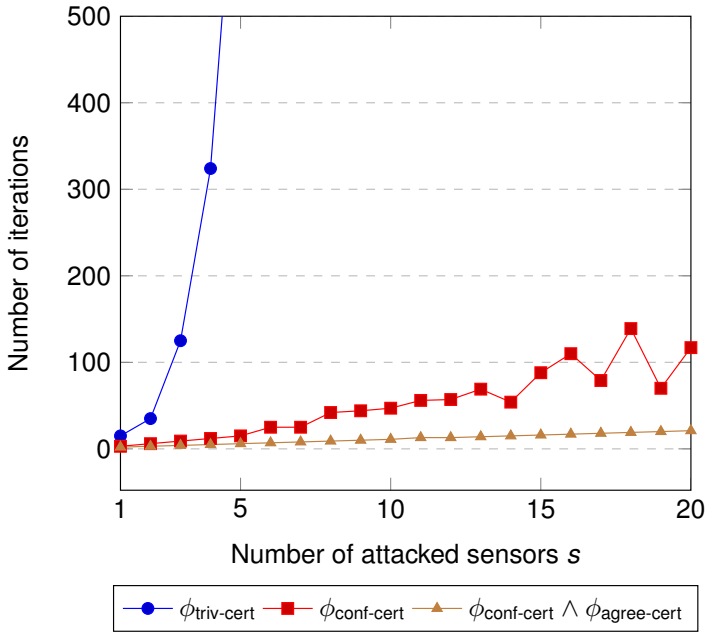
- For a fixed system ($n = 25, p = 60, s = 20$), increase the number of attacked sensors.



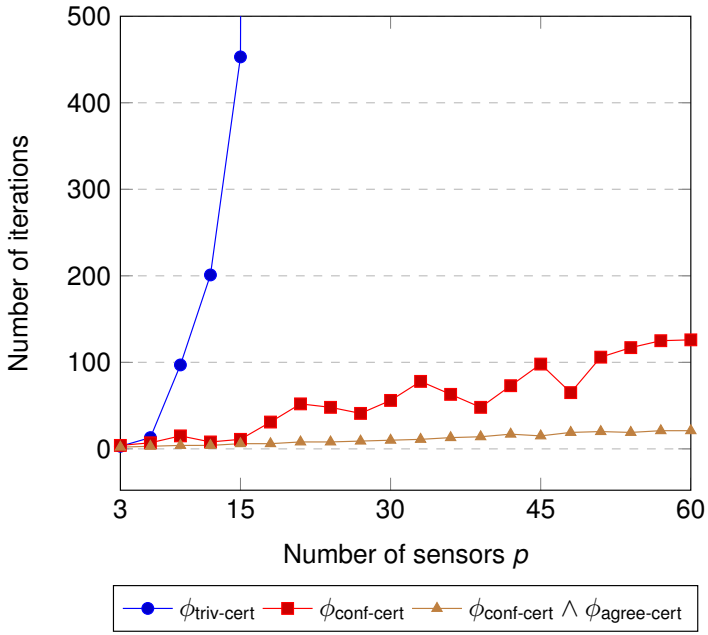
State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Simulation results

- For a fixed system ($n = 25, p = 60, s = 20$), increase the number of attacked sensors.



- Increase the number of sensors from 3 to 60. One third of sensors is under attack.



CyPhyLab UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Simulation results

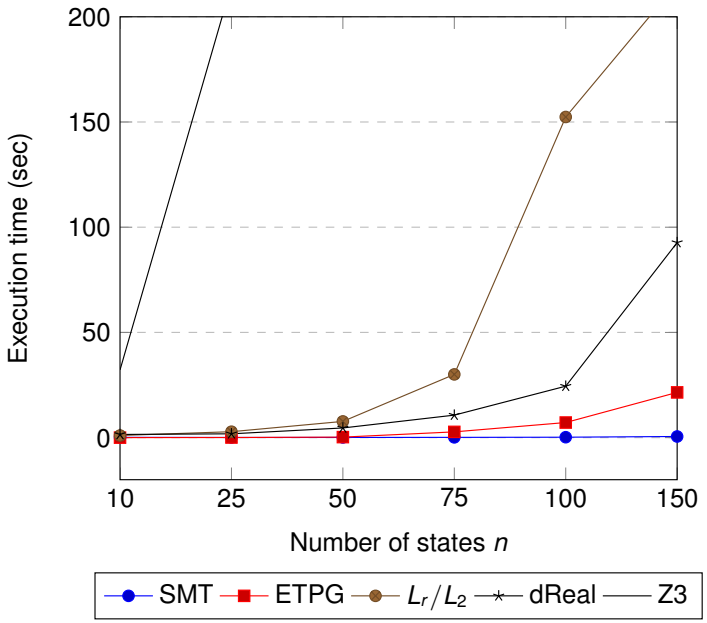
- Comparison against 2 convex-relaxation algorithms and 2 logic-based encodings.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Simulation results

- Comparison against 2 convex-relaxation algorithms and 2 logic-based encodings.

- Fix the number of sensors $p = 60, s = 20$ and increase the number of system states from 10 to 150.

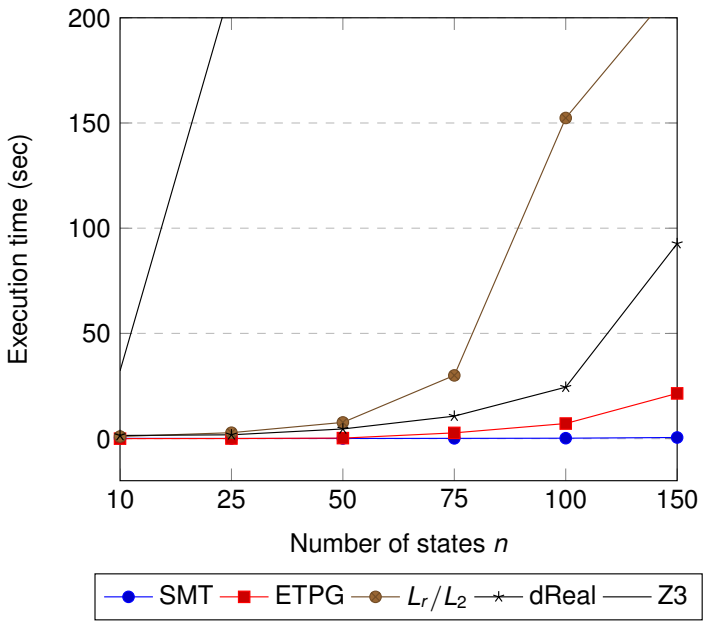


State reconstruction under sensor attacks

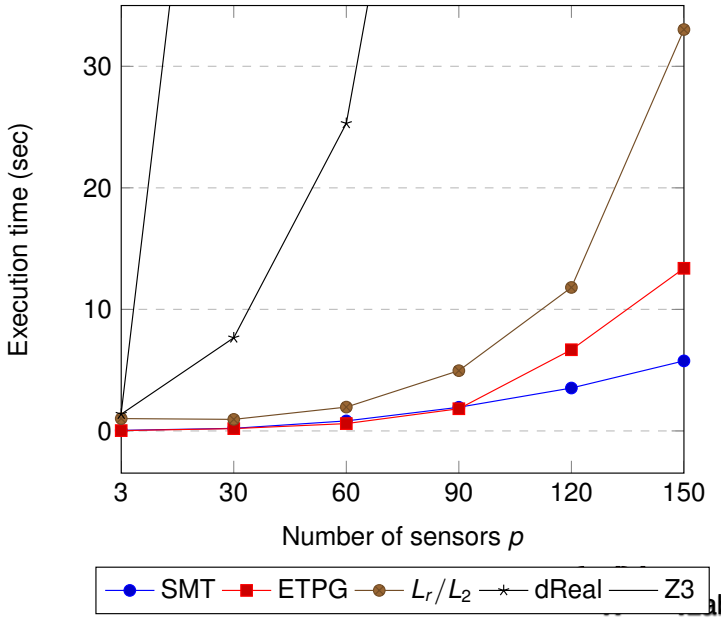
Satisfiability Modulo Theories approach: Simulation results

- Comparison against 2 convex-relaxation algorithms and 2 logic-based encodings.

- Fix the number of sensors $p = 60, s = 20$ and increase the number of system states from 10 to 150.



- Fix the number of states $n = 50$ and increase the number of sensors from 3 to 150 ($s = p/3$).



UCLA

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: ACC example

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Some extensions

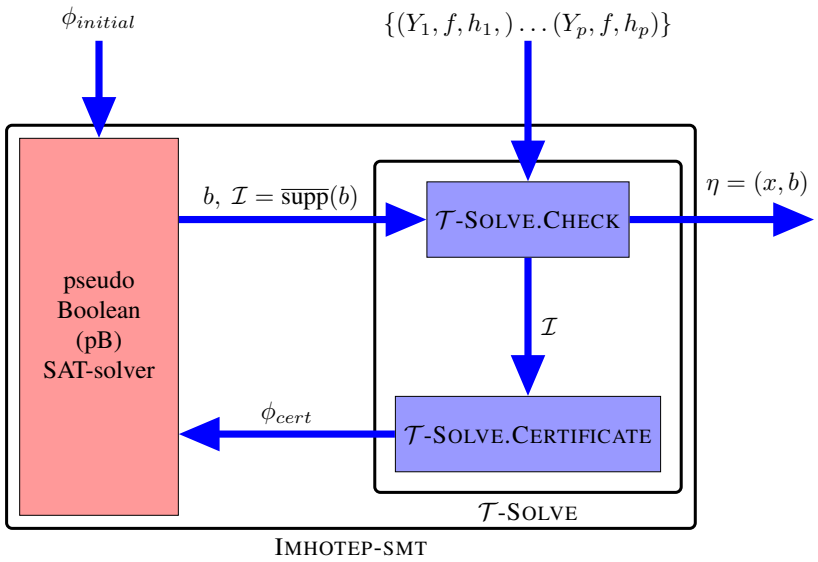
- Bounded noise: batch data.
- Bounded noise: recursive (observer) algorithm.
- Stochastic noise: combine Kalman filters with SMT solving.
- Nonlinear systems: differential flatness.

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Differentially flat systems

System Dynamics:

$$\Sigma_a \begin{cases} x(t+1) & = f(x(t), u(t)), \\ y(t) & = h(x(t)) + a(t) \end{cases}$$



Theorem

Let the nonlinear dynamical system Σ_a be 2s-sparse flat. Then, IMHOTEP-SMT *terminates* ☺ *with:*

- $supp(b^*) \subseteq supp(b)$. ☺.
- $\|x^* - x\|_2^2 = 0$. ☺

Moreover, the upper bound on the number of iterations is $\binom{p}{p-2s+1}$. ☺

State reconstruction under sensor attacks

Satisfiability Modulo Theories approach: Differentially flat systems

Acknowledgments

- Students and collaborators at UCLA, Berkeley, and UPenn;
- National Science Foundation and DARPA;
- Alberto Bemporad and Maurice Heemels for inviting me.

For more information:

Special issue on CPS Security
Control Systems Magazine, 35(1), 2015.

Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks
H. Fawzi, P. Tabuada, and S. Diggavi
IEEE Transactions on Automatic Control, 59(6), 2014.

Event-Triggered State Observers for Sparse Noise/Attacks
Y. Shoukry and P. Tabuada
arXiv:1309.3511, 2014.

Sound and Complete State Estimation for Linear Dynamical Systems under Sensor Attacks
Using Satisfiability Modulo Theory Solving
Y. Shoukry, A. Puggelli, A. Sangiovanni-Vincentelli, S. Seshia, and P. Tabuada
2015 American Control Conference.

Optimal Guarantees Against Sensor Attacks in the Presence of Noise
S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada
arXiv, 2015.

`http://www.cyphylab.ee.ucla.edu/`
`http://www.ee.ucla.edu/~tabuada`

CyPhyLab UCLA