Secure State-estimation and Control of Cyber-Physical Systems

Paulo Tabuada

Cyber-Physical Systems Laboratory
Department of Electrical Engineering
University of California at Los Angeles
Secure State-estimation and Control of Cyber-Physical Systems

Yasser Shoukry

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

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
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.
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-
Cyber-physical systems’ security in the news

**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-
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.
Cyber-physical systems’ security in the news

U.S. Indicts 7 Iranians in Cyberattacks on Banks and a Dam

By DAVID E. SANGER  MARCH 24, 2016

Cyberattackers attempted to gain control of the Bowman Dam in Rye, a suburb of New York, in 2013. The effort failed, but worried American investigators because it was aimed at seizing a piece of infrastructure.  Christopher Capozziello for The New York Times
Why security for cyber-physical systems?

- Cyber-physical systems are **physical** processes (chemical plants, power grids, aircraft, etc.) whose functionality requires a tight integration with **cyber** components (computation and communication hardware/software).

- Cyber-physical systems are becoming increasingly larger, distributed, and open to the cyber-world (e.g., internet): **increased vulnerability to attacks**.
Why security for cyber-physical systems?

- Cyber-physical systems are physical processes (chemical plants, power grids, aircraft, etc.) whose functionality requires a tight integration with cyber components (computation and communication hardware/software).

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

- Attacks launched in the cyber domain can have physical consequences.
Why security for cyber-physical systems?

- Cyber-physical systems are **physical** processes (chemical plants, power grids, aircraft, etc.) whose functionality requires a tight integration with **cyber** components (computation and communication hardware/software).

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

- Attacks launched in the **cyber** domain can have **physical** consequences.

- How to detect, identify, and operate in the presence of **attacks**?
The setup

- Physical process modeled as a linear dynamical system:

\[ x(t + 1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]
The setup

- Physical process modeled as a linear dynamical system:

  \[ x(t + 1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]

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

  \[ y(t) = Cx(t) \]
The setup

- Physical process modeled as a linear dynamical system:
  \[ x(t + 1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]

- A total of \( p \) sensors monitor state of plant \((y(t) \in \mathbb{R}^p)\):
  \[ y(t) = Cx(t) + e(t). \]

- Some sensors are attacked:
  - \( e_i(t) \neq 0 \rightarrow \text{sensor } i \text{ is attacked at time } t; \)
The setup

- Physical process modeled as a linear dynamical system:
  \[ x(t + 1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]

- A total of \( p \) sensors monitor state of plant \( y(t) \in \mathbb{R}^p \):
  \[ y(t) = Cx(t) + e(t). \]

- Some sensors are attacked:
  - \( e_i(t) \neq 0 \rightarrow \) 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.).
The setup

- Physical process modeled as a linear dynamical system:
  \[ x(t+1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]

- A total of \( p \) sensors monitor state of plant (\( y(t) \in \mathbb{R}^p \)):
  \[ y(t) = Cx(t) + e(t). \]

- Some sensors are attacked:
  - \( e_i(t) \neq 0 \) \( \rightarrow \) 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 \( K \subset \{1, \ldots, p\} \):
  \[ \text{supp}(e) = \{i \in \{1, \ldots, p\} \mid e_i \neq 0\} = K. \]
The setup

- Physical process modeled as a linear dynamical system:
  \[ x(t + 1) = Ax(t) + Bu(t), \quad x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, t \in \mathbb{N}_0. \]

- A total of \( p \) sensors monitor state of plant \((y(t) \in \mathbb{R}^p)\):
  \[ y(t) = Cx(t) + e(t). \]

- Some sensors are attacked:
  - \( e_i(t) \neq 0 \) \( \rightarrow \) 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 \( q \): \( q = |K| \);
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 \((y(t) \in \mathbb{R}^p)\):

\[ y(t) = Cx(t) + e(t) \]

- The objective is to design a controller:

\[ u(t) = \phi(t, y(0), \ldots, y(t)) \]

rendering the closed-loop system exponentially stable
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 \( (y(t) \in \mathbb{R}^p) \):
  \[ y(t) = Cx(t) + e(t) \]

- The objective is to design a controller:
  \[ u(t) = \phi(t, y(0), \ldots, y(t)) \]

rendering the closed-loop system exponentially stable, notwithstanding any adversarial attack to \( q \) sensors, i.e., for all \( e(t) \in \mathbb{R}^p \) with support \( K, |K| = q \).
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 \( (y(t) \in \mathbb{R}^p) \):
  \[ y(t) = Cx(t) + e(t) \]

- 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 \( (y(t) \in \mathbb{R}^p) \):
  \[ y(t) = Cx(t) + e(t) \]
  \[ e(t) \] is the attack 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 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 \((y(t) \in \mathbb{R}^p)\):
  \[ y(t) = Cx(t) + e(t) \]

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 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 \( y(t) \in \mathbb{R}^p \):
  \[ y(t) = Cx(t) + e(t) \]
  with support in \( K \) at time \( t \in \mathbb{N}_0 \);

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 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 \( (y(t) \in \mathbb{R}^p) \):
  \[ y(t) = Cx(t) + e(t) \]

- 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 stabilize the plant while the attacker seeks to prevent stabilization.
Questioning the setup

- Are physical systems really linear?
Questioning the setup

- Are physical systems really linear?
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.
Questioning the setup

- Are physical systems really linear?
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.
- Why is the set $K$ of attacked sensors fixed throughout the game?
Questioning the setup

- **Are physical systems really linear?**
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.

- **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

- Are physical systems really linear?
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.

- 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?
  - Our results are independent of where and how the attack is conducted. Can you not protect the sensors or the communication using cyber-security techniques?
Questioning the setup

- Are physical systems really linear?
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.

- 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?
  - Our results are independent of where and how the attack is conducted.
Questioning the setup

- Are physical systems really linear?
  - No! Our first results used ideas from compressed sensing and error correction over the reals, hence linearity. The current understanding allows for nonlinear systems, conceptually.

- 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?
  - Our results are independent of where and how the attack is conducted.

- Can you not protect the sensors or the communication using cyber-security techniques?
Attacking sensors

Wheel Speed Sensor
Tone Wheel
Attacking sensors

Noninvasive spoofing attacks for Anti-Lock Braking systems
Attacking sensors

Noninvasive spoofing attacks for Anti-Lock Braking systems
Attacking sensors

Noninvasive spoofing attacks for Anti-Lock Braking systems
Attacking sensors

Noninvasive spoofing attacks for Anti-Lock Braking systems
A separation result

- The attacks are arbitrary, in particular they can be nonlinear and time-varying.
- Do we need to design a nonlinear and time-varying controller to be resilient to attacks?
A separation result

- The attacks are arbitrary, in particular they can be nonlinear and time-varying.
- Do we need to design a nonlinear and time-varying controller to be resilient to attacks?

**Theorem**

Consider the linear control system:

\[
\begin{align*}
    x(t+1) &= Ax(t) + Bu(t) \\
    y(t) &= Cx(t) + e(t).
\end{align*}
\]

If there exists a controller \( u(t) = \phi(t, y(0), \ldots, y(t)) \) rendering the closed-loop system exponentially stable\(^a\) despite an adversarial attack to \( q \) sensors then there exists a decoder \( D : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^n \) that correctly reconstructs the state in \( n \) steps:

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

\(^a\) for a rate of decay smaller than the smallest eigenvalue of \( A \).
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\(^a\) despite an adversarial attack to \( q \) sensors then there exists a decoder \( D : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^n \) that correctly reconstructs the state in \( n \) steps:

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

\(^a\)for a rate of decay smaller than the smallest eigenvalue of \( A \).

We can solve our initial problem in two steps:

1. design the decoder (observer) \( D \);
2. design a linear static controller.
Error correction

\[ x(t + 1) = Ax(t) + Bu(t) \]
\[ y(t) = Cx(t) + e(t) \]
Error correction

\[ x(t + 1) = Ax(t) + Bu(t) \]
\[ y(t) = Cx(t) + e(t) \]

We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);
Error correction

\[
x(t+1) = Ax(t) \\
y(t) = Cx(t) + e(t)
\]

- We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);
Error correction

\[ x(t + 1) = Ax(t) \]
\[ y(t) = Cx(t) + e(t) \]

- We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);
- A decoder (observer) \( D \) processes observations \( y(0), \ldots, y(T - 1) \) and produces an estimate of the initial state \( x(0) \).
Error correction

\[ x(t + 1) = Ax(t) \]
\[ y(t) = Cx(t) + e(t) \]

- We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);
- A decoder (observer) \( D \) 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:

\[ D(y(0), \ldots, y(T - 1)) = x(0). \]
Error correction

\[
\begin{align*}
x(t + 1) &= Ax(t) \\
y(t) &= Cx(t) + e(t)
\end{align*}
\]

- We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);

- A decoder (observer) \( D \) processes observations \( y(0), \ldots, y(T - 1) \) and produces an estimate of the initial state \( x(0) \).

- We say that a decoder \( D : (\mathbb{R}^p)^T \rightarrow \mathbb{R}^n \) 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.
Error correction

\[ x(t + 1) = Ax(t) \]
\[ y(t) = Cx(t) + e(t) \]

- We assume the input to be known since we design the controller. For simplicity we will take \( u(t) = 0 \) for all \( t \in \mathbb{N}_0 \);

- A **decoder (observer)** \( D \) processes observations \( y(0), \ldots, y(T - 1) \) and produces an estimate of the initial state \( x(0) \).

- We say that a decoder \( D : (\mathbb{R}^p)^T \rightarrow \mathbb{R}^n \) **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.
Correction of \( q \) errors

Necessary and sufficient conditions

- 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.

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.

Information theoretic interpretation: if a pair \((A, C)\) is \( \theta \)-sparse observable, the Hamming distance between a sequence of outputs is at least \( p - \theta + 1 \).

Can we efficiently check sparse observability?

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 \( |\text{supp}(Cv)| > 2q \) for every eigenvector \( v \) of \( A \).
Correction of $q$ errors
Necessary and sufficient conditions

- 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.*
Correction of $q$ errors

Necessary and sufficient conditions

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.

No more than $p/2$ errors can be corrected since $2q$ is necessarily smaller than $p$. 
Correction of $q$ errors

Necessary and sufficient conditions

- 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.

- 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.
Correction of $q$ errors

Necessary and sufficient conditions

- 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.*

- 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.
- **Information theoretic interpretation:** if a pair $(A, C)$ is $\theta$-sparse observable, the Hamming distance between a sequence of outputs is at least $p - \theta + 1$. 

Paulo Tabuada (CyPhyLab - UCLA)
Securing CPS
SoSCYPS 04/11/2016 13 / 31
Correction of \( q \) errors

Necessary and sufficient conditions

- 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.*

- 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.
- Information theoretic interpretation: if a pair \((A, C)\) is \( \theta \)-sparse observable, the Hamming distance between a sequence of outputs is at least \( p - \theta + 1 \).
- Can we efficiently check sparse observability?
Correction of $q$ errors

Necessary and sufficient conditions

- 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.

- 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.
- Information theoretic interpretation: if a pair $(A, C)$ is $\theta$-sparse observable, the Hamming distance between a sequence of outputs is at least $p - \theta + 1$.
- Can we efficiently check sparse observability?

**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 $|\text{supp}(Cv)| > 2q$ for every eigenvector $v$ of $A$. 
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

- So far we discussed when we can or cannot reconstruct the state in the presence of attacks.
- Are there efficient algorithms for state reconstruction?

1. 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.

2. Event-Triggered State Observers for Sparse Noise/Attacks
   Y. Shoukry and P. Tabuada

3. IMHOTEP-SMT: A Satisfiability Modulo Theory Solver For Secure State Estimation
   13th International Workshop on Satisfiability Modulo Theories (SMT) 2015.
So far we discussed when we can or cannot reconstruct the state in the presence of attacks.

Are there efficient algorithms for state reconstruction?

Our first approach was to use the $\ell_0$ to $\ell_1$ relaxation popularized in compressed sensing and error correction over the reals\(^1\).

---

\(^1\) 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.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

- So far we discussed when we can or cannot reconstruct the state in the presence of attacks.

- Are there efficient algorithms for state reconstruction?
  - Our first approach was to use the $\ell_0$ to $\ell_1$ relaxation popularized in compressed sensing and error correction over the reals\(^1\).
  - We then developed our own customized optimization algorithm/observer and provided a tight characterization of when it is guaranteed to work\(^2\).

---

\(^1\) 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.

\(^2\) Event-Triggered State Observers for Sparse Noise/Attacks
Y. Shoukry and P. Tabuada
So far we discussed when we can or cannot reconstruct the state in the presence of attacks.

Are there efficient algorithms for state reconstruction?

- Our first approach was to use the $\ell_0$ to $\ell_1$ relaxation popularized in compressed sensing and error correction over the reals\(^1\).
- We then developed our own customized optimization algorithm/observer and provided a tight characterization of when it is guaranteed to work\(^2\).
- We now present an algorithm that always reconstructs the state when it is possible to do so, i.e., under the right sparse observability assumptions\(^3\).

---

1. 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.
2. Event-Triggered State Observers for Sparse Noise/Attacks
   Y. Shoukry and P. Tabuada
3. IMHOTEP-SMT: A Satisfiability Modulo Theory Solver For Secure State Estimation
   13th International Workshop on Satisfiability Modulo Theories (SMT) 2015.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:

\[
\sum_a \begin{cases} 
  x(t + 1) & = Ax(t) \\
  y(t) & = Cx(t) + a(t)
\end{cases}
\]
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:

$$\sum_a \begin{cases} x(t+1) = Ax(t) \\ y(t) \end{cases}$$

Collect $\tau$ measurements:

$$\begin{bmatrix} y_i(t-\tau+1) \\ y_i(t-\tau) \\ \vdots \\ y_i(t) \end{bmatrix} = \begin{bmatrix} C_i \\ C_iA \\ \vdots \\ C_iA^{\tau-1} \end{bmatrix} x + \begin{bmatrix} a_i(t-\tau+1) \\ a_i(t-\tau) \\ \vdots \\ a_i(t) \end{bmatrix}$$
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:

\[ \Sigma_a \begin{cases} \dot{x}(t+1) = Ax(t) \\ y(t) = Cx(t) + a(t) \end{cases} \]

Collect \( \tau \) measurements:

\[ Y_i = \begin{cases} O_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ O_i x & \text{if sensor } i \text{ is attack-free} \end{cases} \]
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:
\[ \Sigma_a \begin{cases} 
  x(t+1) &= Ax(t) \\
  y(t) &= Cx(t) + a(t) 
\end{cases} \]

Collect \( \tau \) measurements:
\[ Y_i = \begin{cases} 
  O_i x + E_i & \text{if sensor } i \text{ is under attack,} \\
  O_i x & \text{if sensor } i \text{ is attack-free} 
\end{cases} \]

- For each individual sensor, we define a binary indicator variable \( b_i \in \mathbb{B} \) by declaring \( b_i = 1 \) when the \( i \)th sensor is under attack and \( b_i = 0 \) otherwise.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:
\[
\Sigma_a \begin{cases} 
  x(t+1) = Ax(t) \\
  y(t) = Cx(t) + a(t) 
\end{cases}
\]

Collect \( \tau \) measurements:
\[
Y_i = \begin{cases} 
  O_i x + E_i & \text{if sensor } i \text{ is under attack,} \\
  O_i x & \text{if sensor } i \text{ is attack-free} 
\end{cases}
\]

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

Problem

(Secure state-estimation) For the linear control system under attack \( \Sigma_a \), construct an estimate \( \eta = (x, b) \in \mathbb{R}^n \times \mathbb{B}^p \) such that \( \eta \models \phi \), i.e., \( \eta \) satisfies \( \phi \), where \( \phi \) is defined as:

\[
\phi := \bigwedge_{i=1}^p \left( \neg b_i \Rightarrow Y_i = O_i x \right) \quad \land \quad \left( \sum_{i=1}^p b_i \leq s \right).
\]
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach

System Dynamics:
\[ \Sigma_a \begin{cases} x(t + 1) = Ax(t) \\ y(t) = Cx(t) + a(t) \end{cases} \]

Collect \( \tau \) measurements:
\[ Y_i = \begin{cases} O_i x + E_i & \text{if sensor } i \text{ is under attack,} \\ O_i x & \text{if sensor } i \text{ is attack-free} \end{cases} \]

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

Problem
(Secure state-estimation) For the linear control system under attack \( \Sigma_a \), construct an estimate \( \eta = (x, b) \in \mathbb{R}^n \times \mathbb{B}^p \) such that \( \eta \models \phi \), i.e., \( \eta \) satisfies \( \phi \), where \( \phi \) is defined as:

\[ \phi ::= \bigwedge_{i=1}^{p} \left( -b_i \Rightarrow \| Y_i - O_i x \|_2^2 \leq 0 \right) \land \left( \sum_{i=1}^{p} b_i \leq s \right). \]
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.
SMT = pB-SAT solver + $T$-Solver.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory 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
A Satisfiability Modulo Theory Approach: Lazy SMT Architecture I

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

- pB-SAT solver: solves the “boolean version” of the problem.
  - Original formula:
    \[
    \phi := \bigwedge_{i=1}^{p} \left( \neg b_i \Rightarrow \| Y_i - O_i x \|_2^2 \leq 0 \right) \\
    \bigwedge \left( \sum_{i \in 1}^{p} b_i \leq s \right).
    \]
State reconstruction under sensor attacks

A Satisfiability Modulo Theory Approach: Lazy SMT Architecture I

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

- pB-SAT solver: solves the "boolean version" of the problem.
  
  Original formula:
  
  \[
  \phi := \bigwedge_{i=1}^{p} \left( \neg b_i \Rightarrow \|Y_i - O_i x\|_2^2 \leq 0 \right) \\
  \bigwedge \left( \sum_{i=1}^{p} b_i \leq s \right).
  \]

- 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**

A Satisfiability Modulo Theory Approach: Lazy SMT Architecture I

- SMT = pB-SAT solver + \( T \)-Solver.
- pB-SAT solver: solves the “boolean version” of the problem.
  - Original formula:
    
    \[
    \phi ::= \bigwedge_{i=1}^{p} \left( \neg b_i \Rightarrow \| Y_i - \mathcal{O}_i x \|_2^2 \leq 0 \right) \\
    \wedge \left( \sum_{i=1}^{p} b_i \leq s \right).
    \]
  - Replace non-boolean variables with boolean ones
    
    \[
    \phi_{\text{initial}} ::= \bigwedge_{i=1}^{p} \left( \neg b_i \Rightarrow c_i \right) \wedge \left( \sum_{i=1}^{p} b_i \leq s \right)
    \]
  - Pass \( \phi_{\text{initial}} \) to the SAT solver.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Lazy SMT Architecture II

- pB-SAT solver returns an assignment for the variable $b$.
- We extract which sensors are “hypothesized” to be attack free $I$. 

![Diagram](image_url)
pB-SAT solver returns an assignment for the variable $b$.

We extract which sensors are “hypothesized” to be attack free $I$.

Check this assignment.
- pB-SAT solver returns an assignment for the variable $b$.
- We extract which sensors are “hypothesized” to be attack free $I$.
- Check this assignment.

1: **Solve:**

$$x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_I - O_I x \|_2^2$$
pB-SAT solver returns an assignment for the variable $b$.

We extract which sensors are “hypothesized” to be attack free $I$.

Check this assignment.

1: **Solve:**

   $x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_{I} - O_{I} x \|_2^2$

2: **if** $\| Y_{I} - O_{I} x \|_2^2 = 0$ **then**

3: status = SAT; 

6: **end if**

7: return (status, $x$);
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Lazy SMT Architecture II

- pB-SAT solver returns an assignment for the variable $b$.
- We extract which sensors are “hypothesized” to be attack free $I$.
- Check this assignment.

1: **Solve:**
   
   $x := \text{argmin}_{x \in \mathbb{R}^n} \| Y_I - O_I x \|_2^2$

2: if $\| Y_I - O_I x \|_2^2 = 0$ then

3: status = SAT;

4: else

5: status = UNSAT;

6: end if

7: return (status, $x$);
Generate “theory lemma”, “counter example”, or “UNSAT certificate”.

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

Add this “certificate” to the original constraints:

\[
\phi := \phi_{\text{initial}} \land \phi_{\text{triv-cert}}
\]

\[
\text{pseudo Boolean (pB) SAT-solver}
\]

\[
\text{IMHOTEP-SMT}
\]

\[
\text{T-SOLVE.CHECK}
\]

\[
\text{T-SOLVE.CERTIFICATE}
\]

\[
\text{T-SOLVE}
\]

\[
\eta = (x, b)
\]

\[
\{(Y_1, O_1), \ldots, (Y_p, O_p)\}
\]
Generate “theory lemma”, “counter example”, or “UNSAT certificate”.

\[ \phi_{\text{triv-cert}} = \sum_{i \in I} b_i \geq 1 \]
Generate “theory lemma”, “counter example”, or “UNSAT certificate”.

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

Add this “certificate” to the original constraints:

$$\phi := \phi_{\text{initial}} \land \phi_{\text{triv-cert}}$$
Generate “theory lemma”, “counter example”, or “UNSAT certificate”.

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

Add this “certificate” to the original constraints:

\[ \phi := \phi_{\text{initial}} \land \phi_{\text{triv-cert}} \]
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Termination and performance

System Dynamics:

\[
\Sigma_a \begin{cases} 
  x(t+1) = Ax(t) \\
  y(t) = Cx(t) + a(t)
\end{cases}
\]

**Proposition**

*Let the linear dynamical system \( \Sigma_a \) be 2s-sparse observable. Then, IMHOTEP-SMT:

- terminates,
- identifies the attacked sensors,
- and reconstructs the state.*

Moreover, the number of iterations is upper bounded by \( \sum_{s=0}^{S} \binom{p}{s} \).
Why is performance bad?

\[ \phi_{\text{triv-cert}} = \sum_{i \in I} b_i \geq 1 \]
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: UNSAT certificates

- Why is performance bad?
  \[ \phi_{\text{triv-cert}} = \sum_{i \in I} b_i \geq 1 \]

- To enhance performance, we need to generate *compact certificates*. 
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: UNSAT certificates

- Why is performance bad?

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

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

**Lemma**

*Let the linear dynamical system \( \Sigma_a \) be 2s-sparse observable. If \( T\text{-SOLVE.CHECK}(I) \) is UNSAT then there exists a subset \( I \subset \text{supp}(b) \) with \( |I| \leq p - 2s + 1 \) such that \( T\text{-SOLVE.CHECK}(I_{\text{temp}}) \) 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 (observability gramian).
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: UNSAT certificates

Theorem

Let the linear dynamical system $\Sigma_a$ be $2s$-sparse observable. Then, IMHOTEP-SMT:

- **terminates**,  
- **identifies** the attacked sensors,  
- and **reconstructs** the state.

Moreover, the number of iterations is upper bounded by $\binom{p}{p-2s+1}$ (compare to: $\sum_{s'=0}^{s} \binom{p}{s'}$).
Random system with 25 states, 60 sensors and an increasing number of attacked sensors.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Simulation results

Random system with 25 states 60 sensors and an increasing number of attacked sensors.

Random systems with 25 states, 1/3 of sensors under attack, and increasing number of sensors.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Simulation results

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

- Random systems with 60 sensors (20 under attack) and an increasing number of states.

![Graph showing execution time vs. number of states for different algorithms.](image)
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Simulation results

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

- Random systems with 60 sensors (20 under attack) and an increasing number of states.

- Random systems with 50 states, 1/3 of sensors under attack, and increasing number of sensors.
State reconstruction under sensor attacks
A Satisfiability Modulo Theory Approach: Some extensions

- **Stochastic noise:**
  - combine Kalman filters with SMT solving;
  - optimal performance: as good as a minimum mean squared error (MMSE) estimator that knows the attacked sensors\(^1\).

- **Nonlinear systems:** differential flatness and applications to quadcopters\(^2\).

---

1. Secure State Estimation: Optimal Guarantees against Sensor Attacks in the Presence of Noise
   S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, P. Tabuada

2. Secure State Reconstruction in Differentially Flat Systems Under Sensor Attacks Using Satisfiability Modulo Theory Solving
   IEEE Conference on Decision and Control, 2015.
State reconstruction under sensor attacks

Related work

- **Static case, no dynamics:**
  - Detection in Adversarial Environments
    K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo
    IEEE Transactions on Automatic Control, 2014
  - Efficient Computation of a Security Index for False Data Attacks in Power Networks
    IEEE Transactions on Automatic Control, 2014

Paulo Tabuada (CyPhyLab - UCLA)
State reconstruction under sensor attacks

Related work

- **Static case, no dynamics:**
  - Detection in Adversarial Environments
    K. G. Vamvoudakis, J. P. Hespanha, B. Sinopoli, and Y. Mo
    IEEE Transactions on Automatic Control, 2014
  - Efficient Computation of a Security Index for False Data Attacks in Power Networks
    IEEE Transactions on Automatic Control, 2014

- **Unknown input observers:**
  - A Unified Filter for Simultaneous Input and State Estimation of Linear Discrete-time Stochastic Systems
    S. Z. Yonga, M. Zhub, and E. Frazzoli
    Automatica, 2016
  - Delayed Observers for Linear Systems With Unknown Inputs
    S. Sundaram and C. N. Hadjicostis
    IEEE Transactions on Automatic Control, 2007
State reconstruction under sensor attacks

Related work

- Attack detectability:
  - Attack detection and identification in cyber-physical systems
    F. Pasqualetti, F. Doerfler, and F. Bullo
    IEEE Transactions on Automatic Control, 2013
State reconstruction under sensor attacks

Related work

- **Attack detectability:**
  - Attack detection and identification in cyber-physical systems
    F. Pasqualetti, F. Doerfler, and F. Bullo
    IEEE Transactions on Automatic Control, 2013

- **Sparse observability in continuous time:**
  - Observability of linear systems under adversarial attacks
    M. S. Chong, M. Wakaikim and J. P. Hespanha
    American Conference on Cyber-Physical Systems, 2015
State reconstruction under sensor attacks

Related work

- **Attack detectability:**
  - Attack detection and identification in cyber-physical systems
    F. Pasqualetti, F. Doerfler, and F. Bullo
    IEEE Transactions on Automatic Control, 2013

- **Sparse observability in continuous time:**
  - Observability of linear systems under adversarial attacks
    M. S. Chong, M. Wakaikim and J. P. Hespanha
    American Conference on Cyber-Physical Systems, 2015

- **Effect of implementation details (delays, jitter, etc.):**
  - Robustness of Attack-Resilient State Estimators
    M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, G. J. Pappas
    International Conference on Cyber-Physical Systems, 2014
State reconstruction under sensor attacks

Related work

- Asymptotic state reconstruction under attacks:
  - Dynamic State Estimation in the Presence of Compromised Sensory Data
    Y. Nakahira, Y. Mo
    IEEE Conference on Decision and Control, 2015
  - See also Yasser Shoukry’s talk at ICCPS on Wednesday.
State reconstruction under sensor attacks

Related work

- Asymptotic state reconstruction under attacks:
  - Dynamic State Estimation in the Presence of Compromised Sensory Data
    Y. Nakahira, Y. Mo
    IEEE Conference on Decision and Control, 2015
  - See also Yasser Shoukry’s talk at ICCPS on Wednesday.

- Nonlinear systems:
  - Secure State Estimation for Nonlinear Power Systems under Cyber Attacks
    Q. Hu, D. Fooladivanda, Y. H. Chang, C. J. Tomlin
  - Resilient State Estimation against Switching Attacks on Stochastic Cyber-Physical Systems
    S. Z. Yong, M. Zhub, E. Frazzoli
    IEEE Conference on Decision and Control, 2015
Securing Cyber-Physical Systems

Final thoughts

Security for CPS is quite different from cyber-security: there are CPS attacks for which there are no cyber-security defenses; attacks on CPS do not exploit bugs, they exploit features; knowledge of the plant, control architecture, and controller is the key for a successful attack or defense.

Cyber-security is needed for CPS but CPS-security is the last line of defense.

Paulo Tabuada (CyPhyLab - UCLA)
Securing Cyber-Physical Systems

Final thoughts

- Security for CPS is quite different from cyber-security:
  - there are CPS attacks for which there are no cyber-security defenses;
Final thoughts

- Security for CPS is quite different from cyber-security:
  - there are CPS attacks for which there are no cyber-security defenses;
  - attacks on CPS do not exploit bugs, they exploit features;
  - knowledge of the plant, control architecture, and controller is the key for a successful attack or defense.
Final thoughts

- Security for CPS is quite different from cyber-security:
  - there are CPS attacks for which there are no cyber-security defenses;
  - attacks on CPS do not exploit bugs, they exploit features;
  - knowledge of the plant, control architecture, and controller is the key for a successful attack or defense.

- Cyber-security is needed for CPS but CPS-security is the last line of defense.
Acknowledgements

- Students and collaborators;
- National Science Foundation, DARPA, and Northrop Grumman;
- Prof. Mitra and Prof. Dullerud for inviting me.

For more information:
http://www.cyphylab.ee.ucla.edu/
http://www.ee.ucla.edu/~tabuada