Secure Estimation-based Kalman Filter for Cyber-Physical Systems against Adversarial Attacks

Qie Hu*, Young Hwan Chang*, Claire Tomlin

Department of Electrical Engineering and Computer Sciences
UC Berkeley

Cyber Physical Systems on Social Media

SECURITY 6/01/2012 @ 3:59PM | 26,186 views

What Stuxnet's Exposure As An American Weapon Means For Cyberwar

Cyberattack Inflicts Massive Damage on German Steel Factory

POSTED BY: PAUL DECEMBER 21, 2014 14:52 1 COMMENT

Emerging Technology From the arXiv
April 24, 2015

Security Experts Hack Teleoperated Surgical Robot

The first hijacking of a medical telerobot raises important questions over the security of remote surgery, say computer security experts.

Keeping your car safe from hacking

Automakers and NHTSA scramble to protect your privacy and safety

Published: May 07, 2015 06:00 AM

CNN News Video TV Opinions More...
U.S. World Politics Tech Health Entertainment Living Travel Money Sports Watch Live TV

FBI: Hacker claimed to have taken over flight's engine controls

By Evan Perez, CNN
Updated 9:19 PM ET, Mon May 18, 2015

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Cyber Physical Systems on Social Media

How vulnerable are UAVs to cyber attacks?

Kevin G. Coleman, SilverRhino *11:50 a.m. EST February 23, 2015*

DOT and FAA Propose New Rules for Small Unmanned Aircraft Systems

Regulations will facilitate integration of small UAS into U.S. aviation system

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Examples:

Cars are vulnerable to wireless hacking

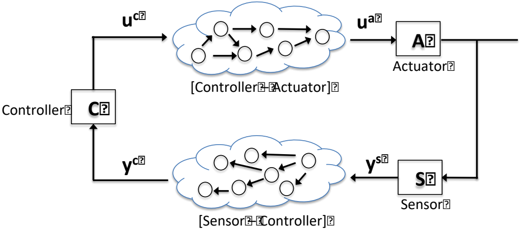
Dan Kaufman (DARPA)
[http://www.cbsnews.com/news/car-hacked-on-60-minutes/]

Rocking Drones with Intentional Sound Noise on Gyro Sensors

[Syssec Lab, KAIST USENIX Security 2015]

Why do we need secure estimation?

- Cyber-physical systems consider **physical systems** controlled by **cyber components** (i.e., communication)
  - Many applications: Power networks, manufacturing processes, air and ground transportation systems



[Conceptual block diagram of a wireless control system]

- Previous work:
  - Attacker's point of view: **[Kosut 2012] [Kwon 2013] [Liu 2011] [Teixeira 2010]**
  - Robust control and filtering methods:
    [Zhou 1998] [Kwon 2013] [Pasqualetti 2011] [Manandhar 2014]
  - Game theory:
    [Roy 2010] [Gupta 2010] [Schwartz 2011][Manshaei 2013] [Gueye 2012][Fei 2013][Amin 2013]

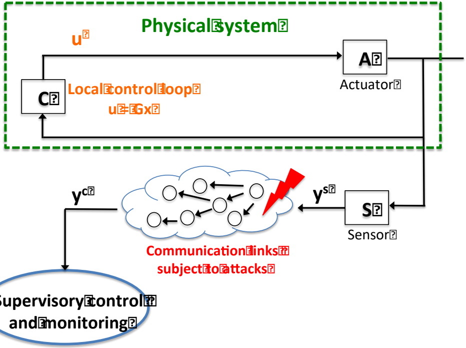
Secure Estimation and Control for CPSs under Adversarial Attacks

Secure Estimation and Control for Cyber-Physical Systems Under Adversarial Attacks

Hamza Fawzi, Paulo Tabuada, *Senior Member, IEEE*, and Suhas Diggavi, *Fellow, IEEE*

Abstract—The vast majority of today's critical infrastructure is supported by numerous feedback control loops and an attack on these control loops can have disastrous consequences. This is a major concern since modern control systems are becoming large and decentralized and thus more vulnerable to attacks. This paper is concerned with the estimation and control of linear systems when some of the sensors or actuators are corrupted by an attacker. We give a new simple characterization of the maximum number of attacks that can be detected and corrected as a function of the pair (A, C) of the system and we show in particular that it is impossible to accurately reconstruct the state of a system if more than half the sensors are attacked. In addition, we show how the design of a secure local control loop can improve the resilience of the system. When the number of attacks is smaller than a threshold, we propose an efficient algorithm inspired from techniques in compressed sensing to estimate the state of the plant despite attacks. We give a theoretical characterization of the performance of this algorithm and we show on numerical simulations that the method is promising and allows to reconstruct the state accurately despite attacks. Finally, we consider the problem of designing output-feedback controllers that stabilize the system despite sensor attacks. We show that a principle of separation between estimation and control holds and that the design of resilient output feedback controllers can be reduced to the design of resilient state estimators.

Index Terms— Algorithm, feedback controller.



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Automatic Control, IEEE Transactions on (Volume:59 , Issue: 6) 2014

Secure Estimation and Control for CPSs under Adversarial Attacks

- Physical process modeled as a linear dynamic system:

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

where some sensors are attacked ($e_i(t) \neq 0$).

- Assumptions:
  - $e_i(t)$ **can be arbitrary** (no stochastic model, no boundedness, etc.)
  - Set of attacked sensors are **fixed**

$$[e(0)\,|\,e(1)\,|\,e(2)\,|\,e(3)] = \begin{bmatrix} 0 & 0 & 0 & 0 \\ * & * & * & * \\ 0 & 0 & 0 & 0 \\ * & * & * & * \end{bmatrix}$$

- Existence of Decoder

[Theorem] There exists a decoder $\mathcal{D}$ that correctly reconstruct the state in *n* steps:

$$x(t-n+1) = \mathcal{D}\big(y(t-n+1), \ldots, y(t)\big)$$

if there exists a controller $u(t)$ rendering the closed-loop system exponentially stable for a sufficient fast rate of decay and despite an adversarial attack to *q* sensors.

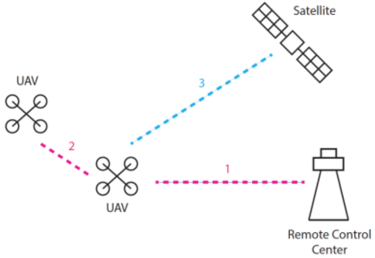
Automatic Control, IEEE Transactions on (Volume:59, Issue: 6) 2014

In this talk,

- Set of **attacked nodes can change over time**
  - Example

$$[e(0)\,|\,e(1)\,|\,e(2)\,|\,e(3)] = \begin{bmatrix} * & 0 & * & 0 \\ 0 & * & 0 & 0 \\ * & * & 0 & 0 \\ 0 & 0 & * & * \end{bmatrix}$$

- Scenarios:
  - Man-In-The-Middle (MITM) Attack in communication with a Remote Control Center or in UAV Formation
  - GPS Spoofing

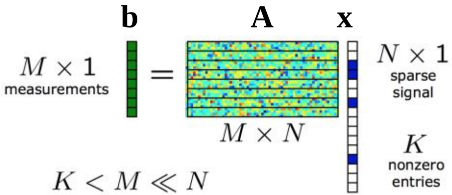


Contributions

- Set of **attacked nodes can change over time:**
  - Showed secure estimation problem in this case is equivalent to classical error correction problem
  - Provided practical method to guarantee existence of accurate decoder
  - Proposed to combine our secure estimator with KF to improve practical performance
  - Simulations of UAVs under adversarial attack

Overview: Compressive Sensing [Candès, Romberg, Tao; Donoho]

- Signal $\mathbf{x}$ is K-sparse
- Replace samples with few linear projections $\mathbf{b} = \mathbf{A}\mathbf{x}$:



- Reconstruction/decoding: given $\mathbf{b} = \mathbf{A}\mathbf{x}$, find $\mathbf{x}$ (L_1 optimization)

> **[Lemma]** If the sparsest solution has $\|x\|_0 = K$ and $\mathbf{M} \geq \mathbf{2}\boldsymbol{K}$ and all subsets of $\mathbf{2}\boldsymbol{K}$ columns of $\mathbf{A}$ are full rank, then the solution is unique.

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Overview: Error Correction [Candès and Tao]

- Wish to transmit n blocks of information $\mathbf{x}$ reliably
- Encode by sending a code word $\mathbf{Cx}$ where $\mathbf{C}$ ($m \times n$ matrix) is a coding matrix ($m \geq n$)
- Assume a fraction of the entries of $\mathbf{Cx}$ are corrupted ➔ $\mathbf{y} = \mathbf{Cx} + \mathbf{e}$
  - Corruption is arbitrary
  - We do not know which entries are corrupted
  - We do not know how the corrupted entries are affected
- Question:
  - Is it possible to recover the signal $\mathbf{x}$ exactly from the corrupted code word?
- Answer:
  - Choose C with random orthonormal columns, then decoding is exact with very high probability

Consider $\mathbf{F}$ where $\mathbf{FC} = \mathbf{0}$:

$$\mathbf{y}' = \mathbf{Fy} = \mathbf{FCx} + \mathbf{Fe} = \mathbf{Fe} \implies \hat{\mathbf{e}} \text{ (applying LP)}$$

$$\hat{\mathbf{x}} = \mathbf{C}^{\dagger}(\mathbf{y} - \hat{\mathbf{e}})$$

Secure Estimation via Error Correction

- Solution:
  - Classical error correction (LP decoding)

- Challenge:
  - To ensure accurate decoding, coding matrix $\mathbf{C}$ must satisfy Restricted Isometry Properties (RIP)
  - Standard solution: construct $\mathbf{C}$ with i.i.d sampled entries
  - In our problem: coding matrix constrained to specific structure

- Questions:
  - What is the connection with Secure Estimation with Fixed Attacked Nodes?
  - Can we provide a more practical way to guarantee the existence of accurate decoder?

Conditions for Existence of Decoder

- Fixed attacked nodes **[H. Fawzi 2014]**:

$$\left|\mathsf{supp}\left(Cz\right) \cup \mathsf{supp}\left(CAz\right) \cup \cdots \cup \mathsf{supp}\left(CA^{T-1}z\right)\right| > 2q, \quad \text{for all } z \in \mathbb{R}^n \setminus \{0\}$$

- Attacked nodes can change over time:

$$\left|\mathsf{supp}\left(\Phi z\right)\right| = \sum_{i=0}^{T-1} \left|\mathsf{supp}\left(CA^i z\right)\right| > 2q \cdot T, \quad \text{for all } z \in \mathbb{R}^n \setminus \{0\}$$

[Lemma] Assume A has n distinct non-zero eigenvalues ($\lambda_i \neq 0$) and T ≥ *n*. Then, the following are equivalent:

(i) $\forall z \in \mathbb{R}^n \setminus \{0\}, |\mathsf{supp}(Cz) \cup \mathsf{supp}(CAz) \cup \cdots \cup \mathsf{supp}(CA^{T-1}z)| > 2q$

(ii) $\forall v_i \in \mathbb{R}^n$ where $Av_i = \lambda_i v_i$ (i.e. eigenvector of A), $|\mathsf{supp}(Cv_i)| > 2q$

(iii) $\forall v_i \in \mathbb{R}^n$ where $Av_i = \lambda_i v_i, |\mathsf{supp}(\Phi v_i)| > 2q \cdot T$

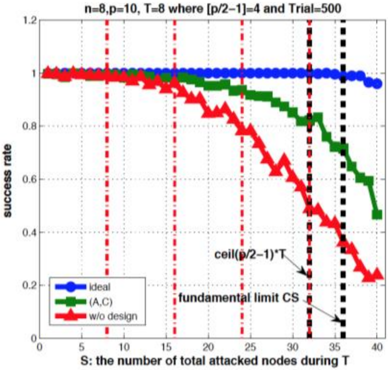
(iv) $\forall z \in \mathbb{R}^n \setminus \{0\}, |\mathsf{supp}(\Phi z)| > 2q \cdot T$

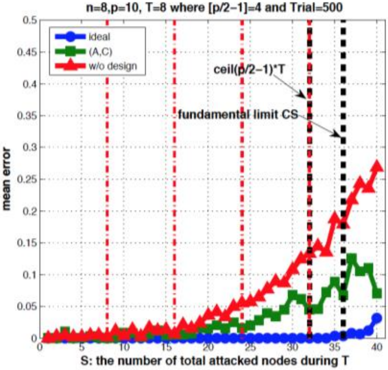
Optimal Decoder Design

[Proposition] Assume that $\mathrm{rank}(\Phi) = n$, the pair (A_o, B) is controllable, and the closed-loop matrix $A(= A_o + BG)$ has n distinct non-zero eigenvalues. Then, the condition for secure estimation of q-errors when the set of attacked nodes is fixed is the same as the condition for when the set of attacked nodes can change over time.

- Optimal decoder design:
  - ❑ Each row of $\mathbf{C}$ is not identically zero
  - ❑ The closed-loop system matrix $\mathbf{A}$ has n distinct non-zero eigenvalues
  - ❑ For all $\boldsymbol{v}_i$ such that $\mathbf{A}\boldsymbol{v}_i = \boldsymbol{\lambda}_i\boldsymbol{v}_i$, $|\mathbf{supp}(\mathbf{C}\boldsymbol{v}_i)| > \mathbf{2q.}$

Numerical Example





Proper design of state feedback gain → Higher success rate, Smaller mean error

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

UAV under Adversarial Attack

- Quadrotor dynamics:

$$x^{(t+1)} = A_0 x^{(t)} + Bu^{(t)} + k + w^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)} + v^{(t)}$$

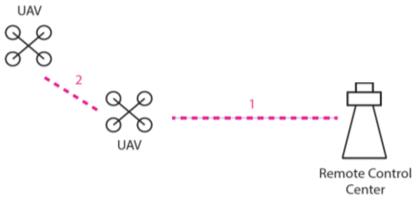
- States: $x = \left[p_x, v_x, \theta_x, \dot{\theta}_x, p_y, v_y, \theta_y, \dot{\theta}_y, p_z, v_z\right]^T$

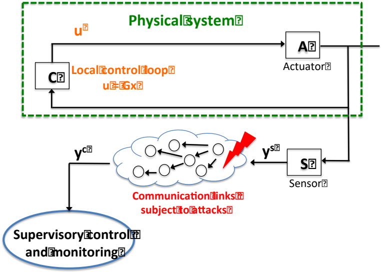
- Controls: $u = \left[\theta_{r,x}, \theta_{r,y}, F\right]^T$

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Example 1: MITM Attack in Communication

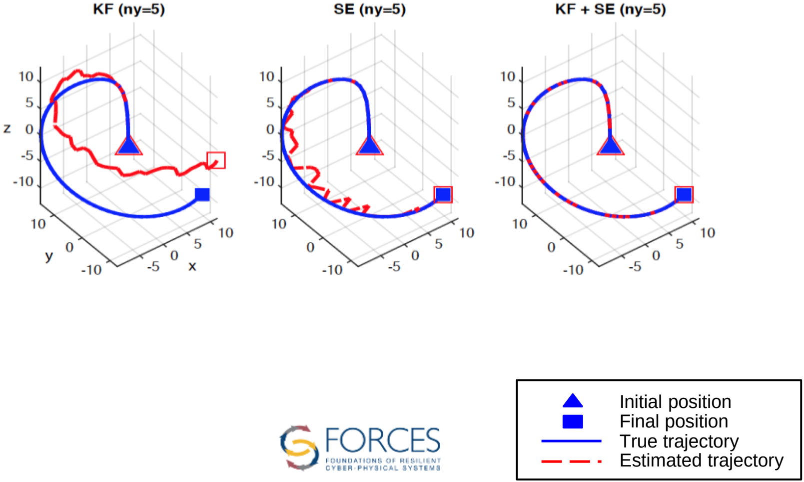
- Setting
  - ❑ Target UAV uses full state feedback
  - ❑ Target UAV sends its position over communication link to a Remote Control Center (RCC) or another UAV in a formation
  - ❑ Communication link attacked, position information corrupted
- Goal
  - ❑ RCC or another UAV to correctly estimate target UAV's trajectory





FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Example 1: MITM Attack in Communication

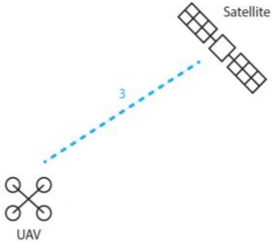


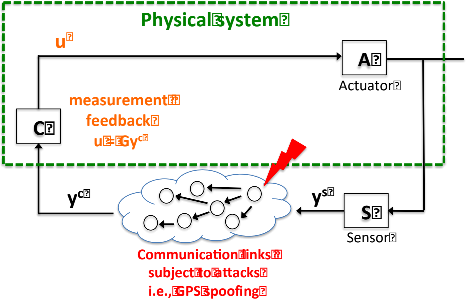
FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Example 2: GPS Spoofing

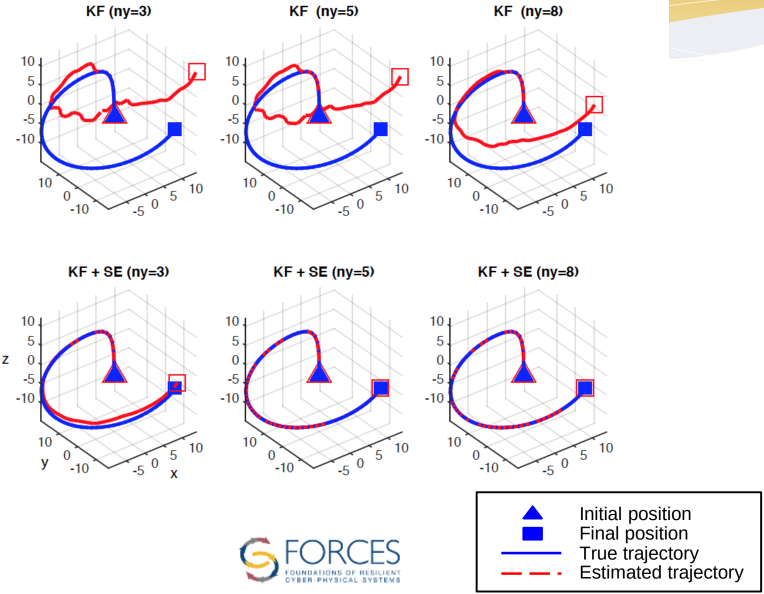
- Setting
  - ❑ Measurement feedback
  - ❑ Position measurements from GPS signal
  - ❑ GPS signal corrupted

- Goal
  - ❑ UAV to follow planned trajectory





Example 2: GPS Spoofing



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

Conclusion

- Secure estimation for CPS under adversarial attacks
- Attack signal:
  - ❑ Set of attack nodes can change from time to time
  - ❑ Arbitrary, does not follow any model
- Proposed secure estimator:
  - ❑ Via error correction tools
  - ❑ Computationally efficient
  - ❑ Outperforms standard KF
- Proposed a practical way to ensure accurate decoding
- Proposed to combine secure estimator with KF for better practical performance
- Simulated UAVs under adversarial attacks
  - ❑ MITM attack
  - ❑ GPS spoofing

Backup slides

Connection between Secure Estimation and Error (T=1)

- Previous work [Fawzi *et al.*, 2014]:

Definition 1. ([1]) *q errors are correctable after T steps by the decoder $\mathcal{D} : (\mathbb{R}^p)^T \to \mathbb{R}^n$ if for any $x^{(0)} \in \mathbb{R}^n$, any $K \subset \{1,..,p\}$ with $|K| \leq q$, and any sequence of vectors $e^{(0)},...,e^{(T-1)}$ in $\mathbb{R}^p$ such that $\mathsf{supp}(e^{(t)}) \subset K$, we have $\mathcal{D}(y^{(0)},...,y^{(T-1)}) = x^{(0)}$ where $y^{(t)} = CA^t x^{(0)} + e^{(t)}$ for $t = 0,...,T-1$.*

Proposition 1. ([1]) *Let $T \in \mathbb{N}\backslash\{0\}$. The following are equivalent:*
(i) There is a decoder that can correct q errors after T steps;
(ii) For all $z \in \mathbb{R}^n\backslash\{0\}$, $|\mathsf{supp}(Cz) \cup \mathsf{supp}(CAz) \cup \cdots \cup \mathsf{supp}(CA^{T-1}z)| > 2q$.

- Classical Error Correction (*T=1*):

Proposition 2. *Consider a $p \times n$ matrix C where $p > n$ and C is full column rank. The following are equivalent for there to exist a q-error-correcting decoder:*
(i) for any $z \neq 0$, $|\mathsf{supp}(Cz)| > 2q$;
(ii) all subsets of $2q$ columns of F are linearly independent where $\mathcal{N}(F) = \mathcal{R}(C)$.

UAV under Adversarial Attack

- We consider a quadrotor with the following dynamics

$$x^{(t+1)} = A_0 x^{(t)} + Bu^{(t)} + k + w^{(t)}$$
$$y^{(t)} = Cx^{(t)} + e^{(t)} + v^{(t)}$$

$$x = [p_x, v_x, \theta_x, \dot{\theta}_x, p_y, v_y, \theta_y, \dot{\theta}_y, p_z, v_z]^T$$

$$A_0 = \begin{bmatrix} 1 & T_s & g\cdot\frac{T_s^2}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & gT_s & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_\theta^{11} & A_\theta^{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & A_\theta^{21} & A_\theta^{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & T_s & g\frac{T_s^2}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & g\cdot T_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_\theta^{11} & A_\theta^{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & A_\theta^{21} & A_\theta^{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & T_s \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ B_\theta^1 & 0 & 0 \\ B_\theta^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & B_\theta^1 & 0 \\ 0 & B_\theta^2 & 0 \\ 0 & 0 & \frac{k_T\cdot T_s^2}{2m} \\ 0 & 0 & \frac{k_T\cdot T_s}{m} \end{bmatrix}$$

$$C_I = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS

UAV under Adversarial Attack

- Quadrotor dynamics:

Process noise

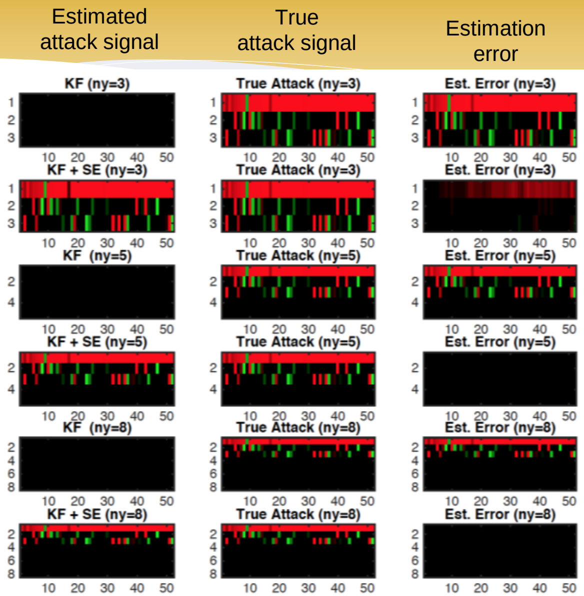
$$
\begin{aligned}
x^{(t+1)} &= A_0 x^{(t)} + Bu^{(t)} + k + \boxed{w^{(t)}} \\
y^{(t)} &= Cx^{(t)} + \boxed{e^{(t)}} + \boxed{v^{(t)}}
\end{aligned}
$$

Attack signal

Measurement noise

$$x = [p_x, v_x, \theta_x, \dot{\theta}_x, p_y, v_y, \theta_y, \dot{\theta}_y, p_z, v_z]^T$$

Example 2: GPS Spoofing



FORCES
FOUNDATIONS OF RESILIENT CYBER-PHYSICAL SYSTEMS