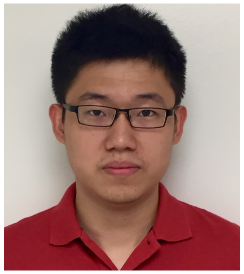
Security of Cyberphysical Systems

P. R. Kumar and Le Xie
Dept. of Electrical and Computer Engineering
Texas A&M University





With Bharadwaj Satchidanandan, Woo Hyun Ko and Tong Huang

Email: prk.tamu@gmail.com
Web: http://cesg.tamu.edu/faculty/p-r-kumar/

Teleseminar
Internet2 CINC UP
September 8, 2017

Securing an automated transportation system

Video "Tackling Autonomous Vehicle Cybersecurity Issues" *at https://cesg.tamu.edu/faculty/p-r-kumar/convergencelab/*

Cyber-physical systems

- Next generation of engineered systems in which computing, communication, and control technologies are tightly integrated
- Many societally important future applications
  - Automated transportation
  - Smart grid
  - Unmanned Air Vehicle Transportation System
  - Water treatment facilities
  - Telesurgery systems
  - …
- Safety critical
  - Malfunctioning causes physical harm
- Critical infrastructure
  - Important to functioning of economy and society

Vulnerability of cyberphysical systems to attacks

- Hackers hitherto could tamper only with information or bits in cyber layer

- CPS tightly couples cyber and physical worlds
  - Actions in physical world taken based on information from cyber layer

- CPS, therefore, gives hacker ability to cause damage in physical world

Security of CPS

- As more systems are connected to the Internet and become more open, there are increasingly more vulnerabilities
- Can be more harmful than other violent attacks
- Next war may be “cyber” rather than “bombs”?
- Even after many decades we still cannot secure the Operating Systems
  - New patches every day
- We still cannot secure the Internet
- Interaction between bits and physical world is very complex
- How can we possibly secure CPSs?

Several attacks on critical infrastructure systems

- Several instances of attacks in the past
  - Maroochy-Shire sewage treatment plant
  - Davis-Besse nuclear power plant
  - Stuxnet
  - Ukraine power grid
  - Water filtering plant in Pennsylvania
  - Demonstrations of cyber attacks in automated cars
- Maroochy-Shire, Australia, 2003, attack on sewage treatment system, commands issued which led to a series of faults in the system
- Attack on computers controlling Davis-Besse nuclear power plant in Ohio, 2003, Slammer worm disabled the safety monitoring system
- Stuxnet worm, 2010, exploited Microsoft Windows vulnerability to subvert critical computers controlling centrifuges in Iran uranium enrichment facility
- Attacks on Supervisory Control and Data Acquisition system, natural gas pipeline systems, trams, power utilities, and water systems, etc.

Isn't network security enough for CPS security?

- Network and information security implemented through periodic patching.
  - CPS has a dynamic system in the loop, and may not admit controllers going online for patching

- Traditional notion of “Confidentiality, Integrity and Availability” in network and information security does not address real-time availability, which is critical for control system security

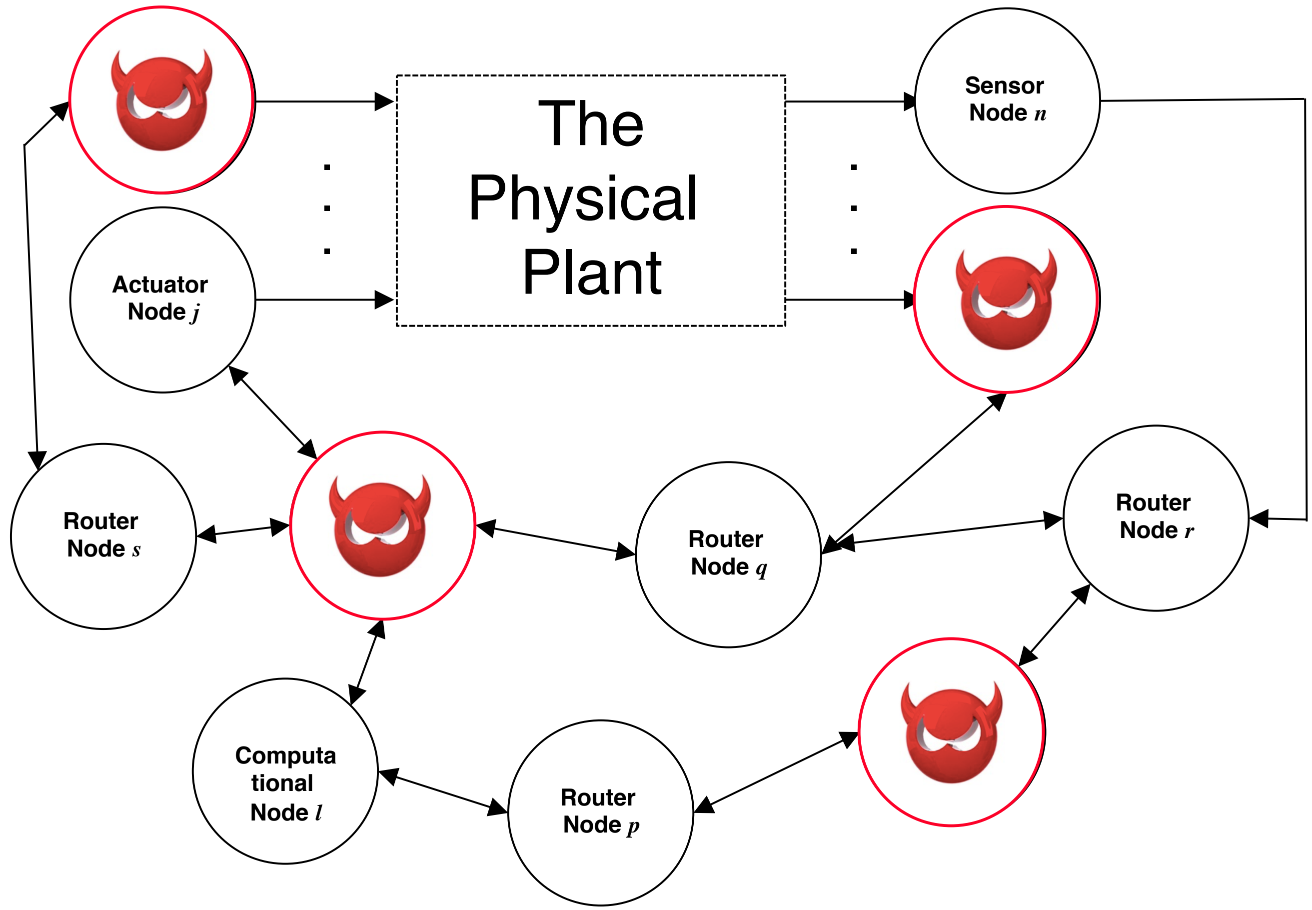
- Network or information security fundamentally cannot address physical layer attacks such as in Maroochy-Shire incident

Two-layer approach to CPS security

- Can think of CPS as consisting of two layers:
  - Cyber layer consisting routers, switches, relays, etc. providing communication backbone,
  - Physical layer consisting the plant, sensors and actuators, controllers which manipulate physical signals
- Cyber layer possibly secured using techniques such as cryptography
  - Therefore, network may possibly be abstracted as secure, reliable, delay-guaranteed bit pipes
- But how to secure the physical layer?

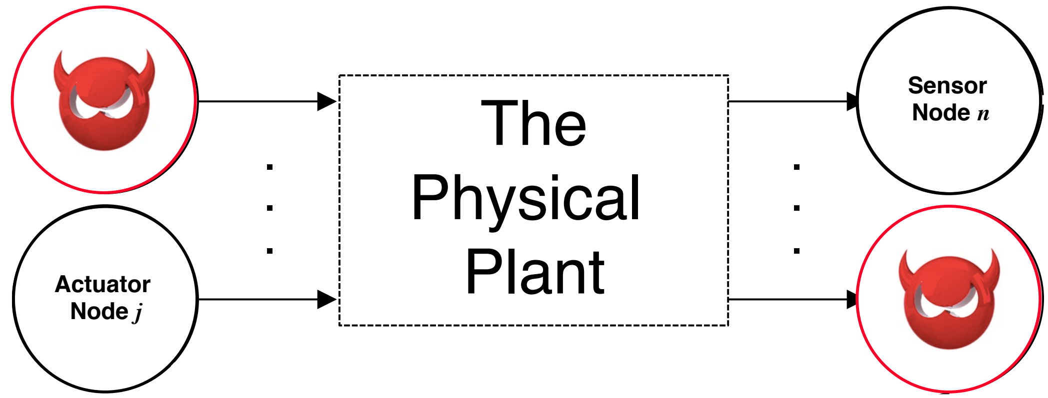
Abstraction of cyberphysical systems

- Overall system has
  - Physical plant
  - Actuators
  - Sensors
  - Routers
  - Computational nodes
  - Network
- But some of the routers, computation nodes, sensors, actuators may be compromised
- How do we secure the overall cyberphysical system?



Abstraction of security problem

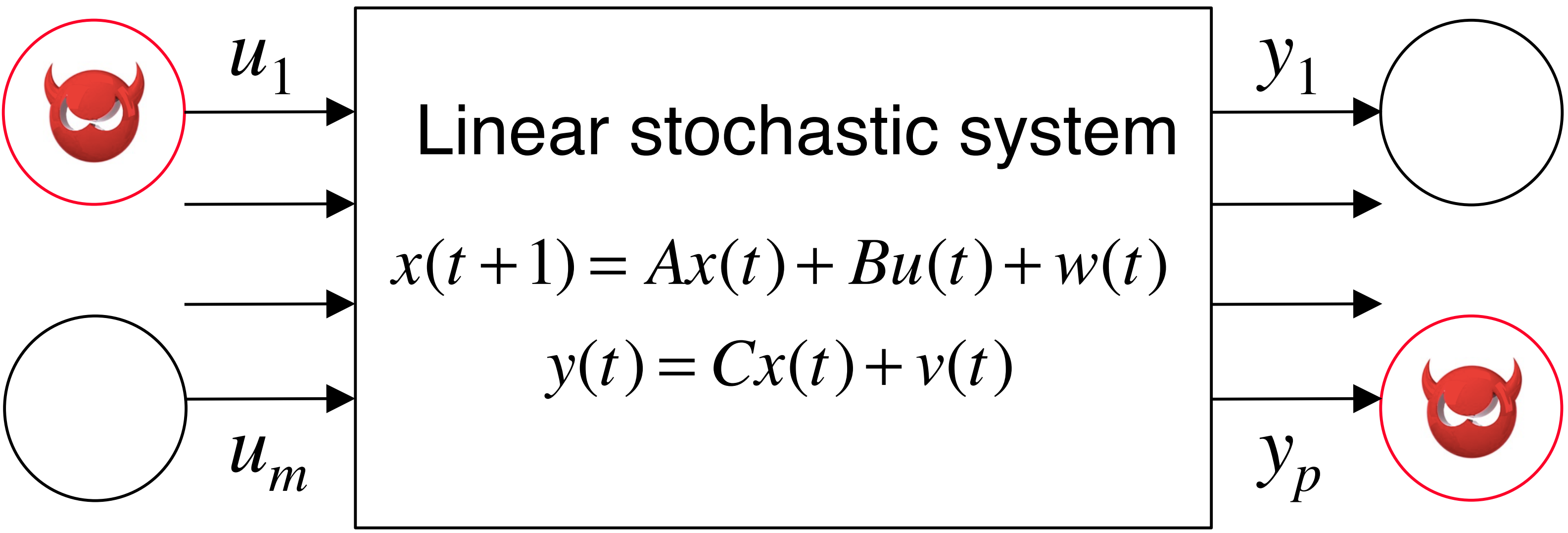
- Some sensors, actuators may be compromised
- If information from a sensor is compromised, we say sensor is compromised
- It does not matter whether sensor is compromised or its information is compromised downstream



How do we secure the overall cyberphysical system when some sensors and actuators may be compromised?

Towards a paranoid theory of linear stochastic systems

Let’s start with linear stochastic systems



Can honest nodes diagnose system?

What performance can they achieve?

- Physical plant modeled as linear stochastic system
  - Most common practical design
- Some actuators/sensors malicious
- Malicious actuators/sensors can collude
- Honest actuators/sensors don't know which nodes malicious

Linear systems theory in a more innocent age

- Linear system $x(t+1) = Ax(t) + Bu(t)$
- When is system controllable (Kalman)?

$$x(n) = A^n x(0) + Bu(n-1) + ABu(n-2) + \ldots + A^{n-1}Bu(0)$$

$$x(n) - A^n x(0) = [B, AB, A^2B, \ldots, A^{n-1}B] \begin{bmatrix} u(n-1) \\ u(n-2) \\ \vdots \\ u(0) \end{bmatrix}$$

- Controllable subspace = Span$[B, AB, \ldots, A^{n-1}B]$
- System is stabilizable if unstable modes of A are in controllable subspace

Linear systems theory in a more innocent age

- Linear system $$x(t+1) = Ax(t)$$ $$y(t) = Cx(t)$$
- When is system state observable from outputs?

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} x(0)$$

- Unobservable subspace = Null Space of $\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$
- System is detectable if unstable modes are observable

But what if some actuators or sensors are malicious?

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \\ \vdots \\ x_n(t+1) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix} \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix}$$

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_p(t+1) \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{p1} & c_{p2} & \dots & c_{pn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}$$

- Some of the u_i's and y_j's may be malicious
- What harm can malicious sensors/actuators cause without the honest sensors/actuators knowledge?

Innocent age concerns vs New age concerns

- Nature causes stability/instability
- Malicious agents cause harm

- Stability of benign systems
- Security of malicious systems

- Stabilizability/Detectability of benign systems
- Securability of malicious systems

Passive guarantees based on system structure

The securable and unsecurable subspaces for deterministic systems

- What states can the malicious sensors/actuators drive the system to without the honest sensors/actuators finding out?

- The unsecurable subspace V is the set of states that the malicious sensors and actuators can keep indistinguishable from the 0 state

- The securable subspace is $V^{\perp}$

The unsecurable states of deterministic systems

- Suppose $x(t+1) = Ax(t) + B_m u_m(t)$

$$y_h(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_H(t) \end{bmatrix} = C_h x(t)$$

- Then $x(0)$ can be made indistinguishable from 0 if for some $u_m(0), u_m(1), \ldots, u_m(t), \ldots$

$$C_h x(0) = 0$$

$$C_h\left(Ax(0) + B_m u_m(0)\right) = 0$$

$$\vdots$$

$$C_h\left(A^t x(0) + A^{t-1} B_m u_m(0) + \ldots B_m u_m(t-1)\right) = 0$$

Characterization of Unsecurable and Securable subspaces

- Unsecurable subspace is the maximal subspace V such that for all v in V

$$C_h v = 0$$

$$\text{There exists } u \text{ such that } Av + B_m u \in V$$

- Securable subspace is $V^{\perp}$

Stochastic systems

Malicious sensors and actuators in linear stochastic system

- Consider a linear stochastic system

$$x(t+1) = Ax(t) + Bu^g(z^t) + B_m u_m(t) + w(t+1)$$
$$y(t) = x(t)$$

- w is white noise of variance Σ
- Honest sensors measure $y_1, y_2, \dots, y_H$
- Malicious sensors measure $y_{H+1}, y_{H+2}, \dots, y_p$
- Sensor measurements reported are $z(t)$, where $z_i(t) = x_i(t)$ for $i = 0, 1, \dots, H$
- But for the malicious sensor's $z_i(t)$ need not equal $x_i(t)$ for $i = H+1, H+2, \dots, p$
- And malicious actuators may apply $u_m(t)$ different from 0

What performance can be guaranteed for a linear stochastic system?

- Honest sensors conduct Test to detect if there is any malicious activity:

$$\lim \frac{1}{T}\sum_{0}^{T-1}\left(z(t+1)-Az(t)+Bu^{g}(z^{t})\right)\left(z(t+1)-Az(t)+Bu^{g}(z^{t})\right)^{T}=\Sigma$$

- To remain undetected malicious sensors/actuators must pass Test

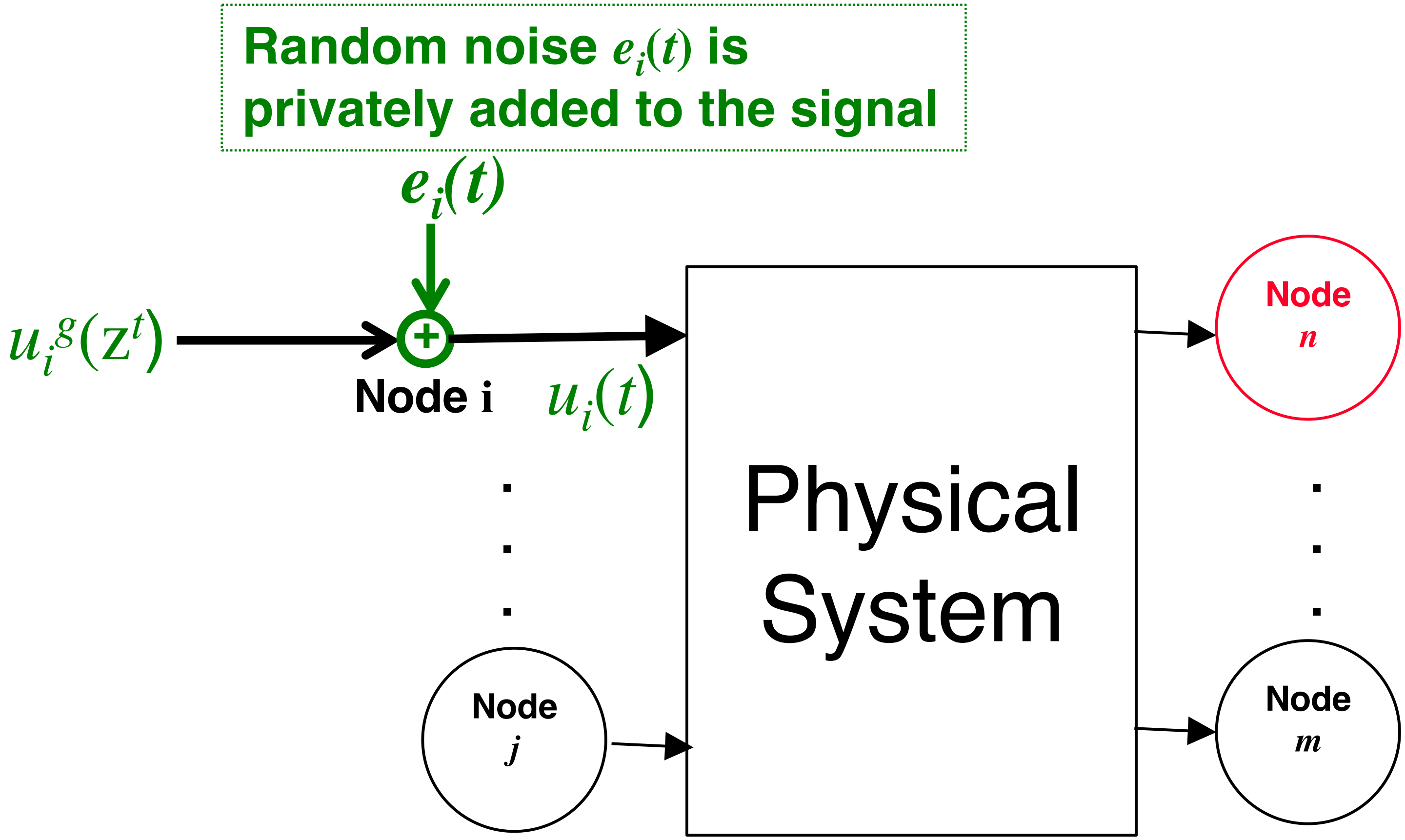
- **Theorem**: Then the error in the reported state error in the securable subspace $V^{\perp}$ is guaranteed to be of zero power

$$\lim \frac{1}{T}\sum_{0}^{T}\left\|\tilde{x}(t)_{V^{\perp}}\right\|^{2}=0$$

$$\text{where } \tilde{x}(t)_{V^{\perp}} := \text{Projection of } \left(z(t)-x(t)\right) \text{ on } V^{\perp}$$

Can we do better?

Dynamic watermarking



- Actuator node superimposes a private excitation whose realization is unknown to other nodes

Why does it help?

- Private excitation $e_i(t)$ appears in transformed returned signals from sensors at time $t+1$
- Measurement reported by sensor at time $t+1$ has to contain suitably transformed contribution of $e_i(t)$
- So actuator can check if private excitation comes back properly from sensors
- Checks if the reported measurements have the appropriately correlations with $e_i(t)$ reported
- This provides powerful guarantees against general attacks on sensors – not just replay attack

Illustration on simple first order SISO system

- SISO system: $x(t+1) = ax(t) + bu(t) + w(t+1)$
  $w(t) \sim N(0, \sigma_w^2)$, i.i.d.
- Dynamic watermarking $u(t) = u^g(t) + e(t)$ with $e(t) \sim N(0, \sigma_e^2)$, i.i.d.
- Two tests are conducted by actuator

$$\lim \frac{1}{T} \sum_{t=0}^{T-1} \left(z(t+1) - az(t) - bu^g(t) - be(t) \right)^2 \stackrel{?}{=} \sigma_w^2$$

$$\lim \frac{1}{T} \sum_{t=0}^{T-1} \left(z(t+1) - az(t) - bu^g(t) \right) \stackrel{?}{=} b^2 \sigma_e^2 + \sigma_w^2$$

- If either test fails, then there is malicious sensor information
  - System goes into safety mode
  - Halted, checked, rebooted, manual operation, etc

Guarantee provided by Dynamic Watermarking

- Theorem

$$\lim \frac{1}{T} \sum_{t=0}^{T-1} v^2(t) = 0$$

- Where $v(t+1) := z(t+1) - az(t) - bu^g(t) - be(t) = w(t+1)$

- **Interpretation**:

$$z(t+1) - az(t) - bu^g(t) - be(t) = w(t+1) + v(t+1)$$

- So reported sensor measurements can distort actual noise $w(t)$ only by zero power signal $v(t)$

Stability consequences of Dynamic Watermarking

- **Theorem**:
- Suppose $|a| < 1$, i.e., system is open-loop stable,
- Then distortion $d[t] := z[t] - x[t]$ is zero power: $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} d^2[k] = 0$
- Mean-square performance is same as reported performance $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] = \lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} z^2[k]$
- Suppose $u^g(t) = fx(t)$ with $|a+bf| < 1$
- Then mean square performance is optimal $\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} x^2[k] = \frac{\sigma_w^2 + b^2\sigma_e^2}{1-|a+bf|^2}$

More general results

- Results extend to

- ARMAX Systems used in process control:

$$y[t] = -\sum_{k=1}^{p} a_k y[t-k] + \sum_{k=0}^{h} b_k u[t-l-k] + \sum_{k=0}^{r} c_k w[t-k]$$

- MIMO partially observed Gaussian systems

$$\mathbf{x}[t+1] = A\mathbf{x}[t] + Bu[t] + \mathbf{w}[t+1]$$
$$y[t+1] = C\mathbf{x}[t+1] + n[t+1]$$

- Some non-Gaussian systems

Example

- System: $y(t+1)+0.7y(t)-0.2y(t-1)=u(t)+0.5u(t-1)+w(t)$

  $w(t)\sim N(0,1)$, i.i.d.

- Actuator applies $u(t)=-0.7z(t)-0.2z(t-1)-0.5u(t-1)+e(t)$

  $e(t)\sim N(0,1)$, i.i.d.

- Closed-loop system:

$$y[t+1]=0.7(y[t]-z[t])+0.3(y[t-1]-z[t-1])+e[t]+w[t+1]$$

- Sensor estimates process noise by

$$\widehat{w}[t+1]:=\frac{1}{2}(y[t+1]-0.7(y[t]-z[t])-0.3(y[t-1]-z[t-1])$$

Example

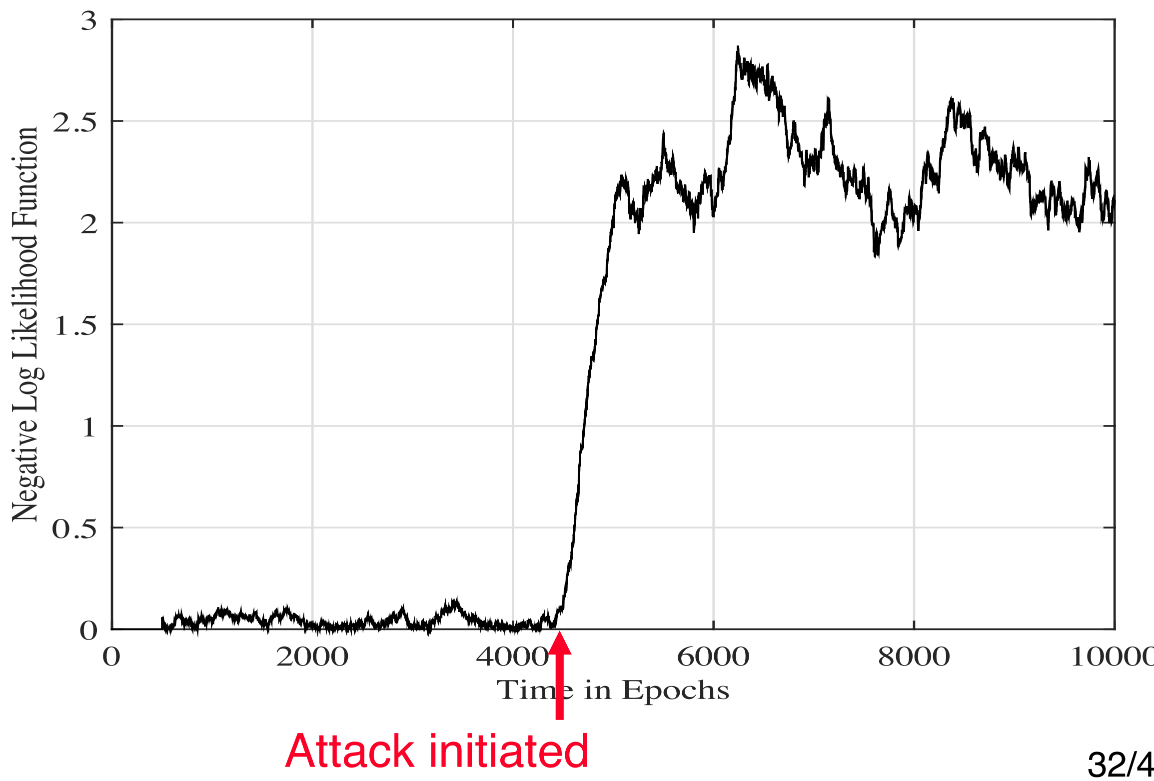
- Simulates a fake system with a fake noise $n(t) - \widehat{w}(t)$

  $n(t) \sim N(0,1)$, i.i.d.

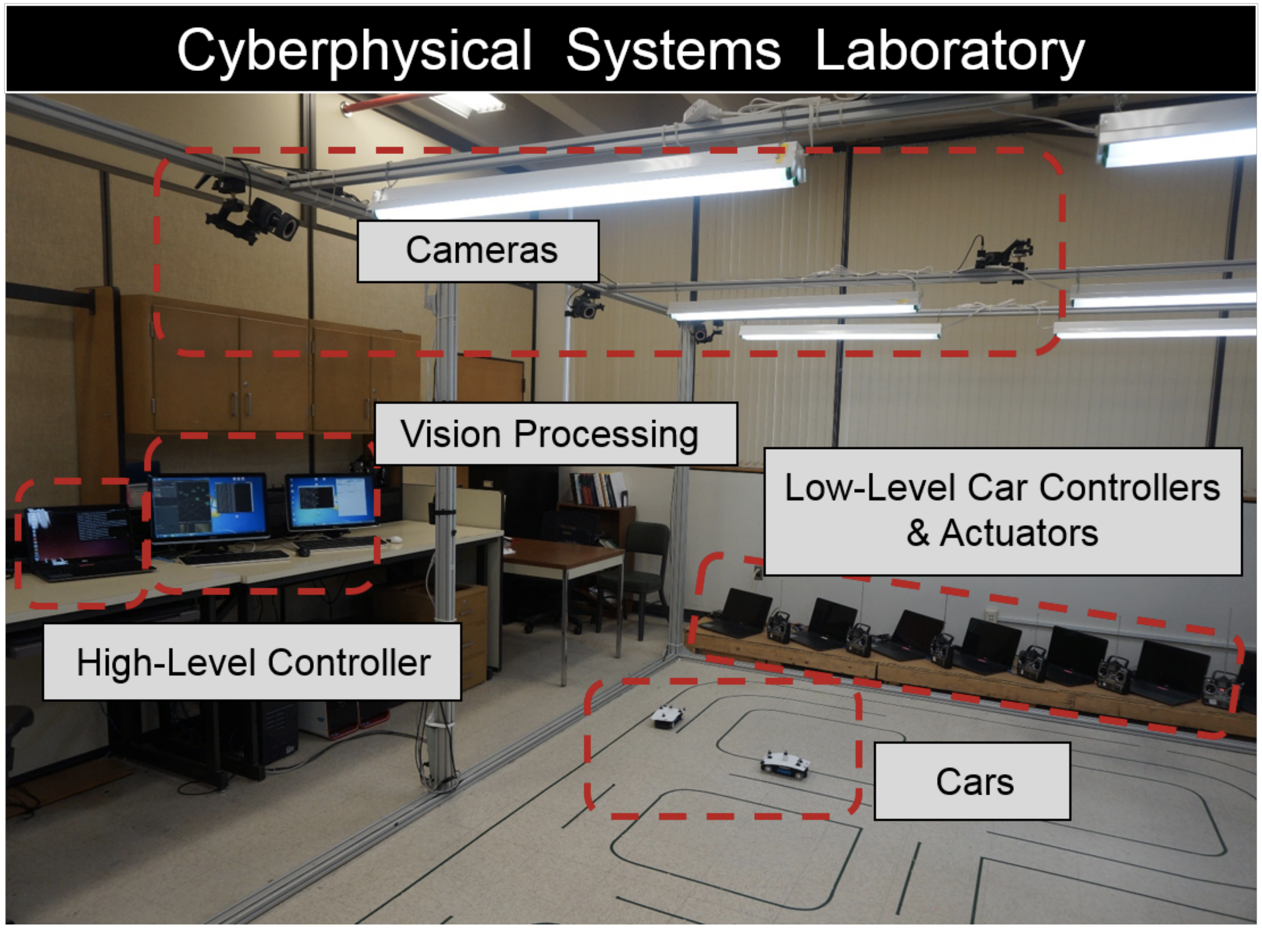
- Reports output of fake simulated system

- In absence of watermarking, actuator would not suspect any malicious measurements

- Sensor attack begins at time 4500



Test of autonomous transportation system in CPS lab



Automated vehicles are vulnerable to cyber attacks

- Hackers have demonstrated remote hijack of a Jeep's digital systems over the Internet
  - Resulted in the car manufacturer recalling over a million units to patch identified security vulnerabilities
- Automated cars use various sensors
  - Ultrasound sensor to determine distance of close objects
  - mm-wave radar to map road immediately ahead
- These sensors can be jammed.
  Researchers from Zhejiang University have demonstrated such sensor attacks
- Several other demonstrations reported recently

Attacks on cars

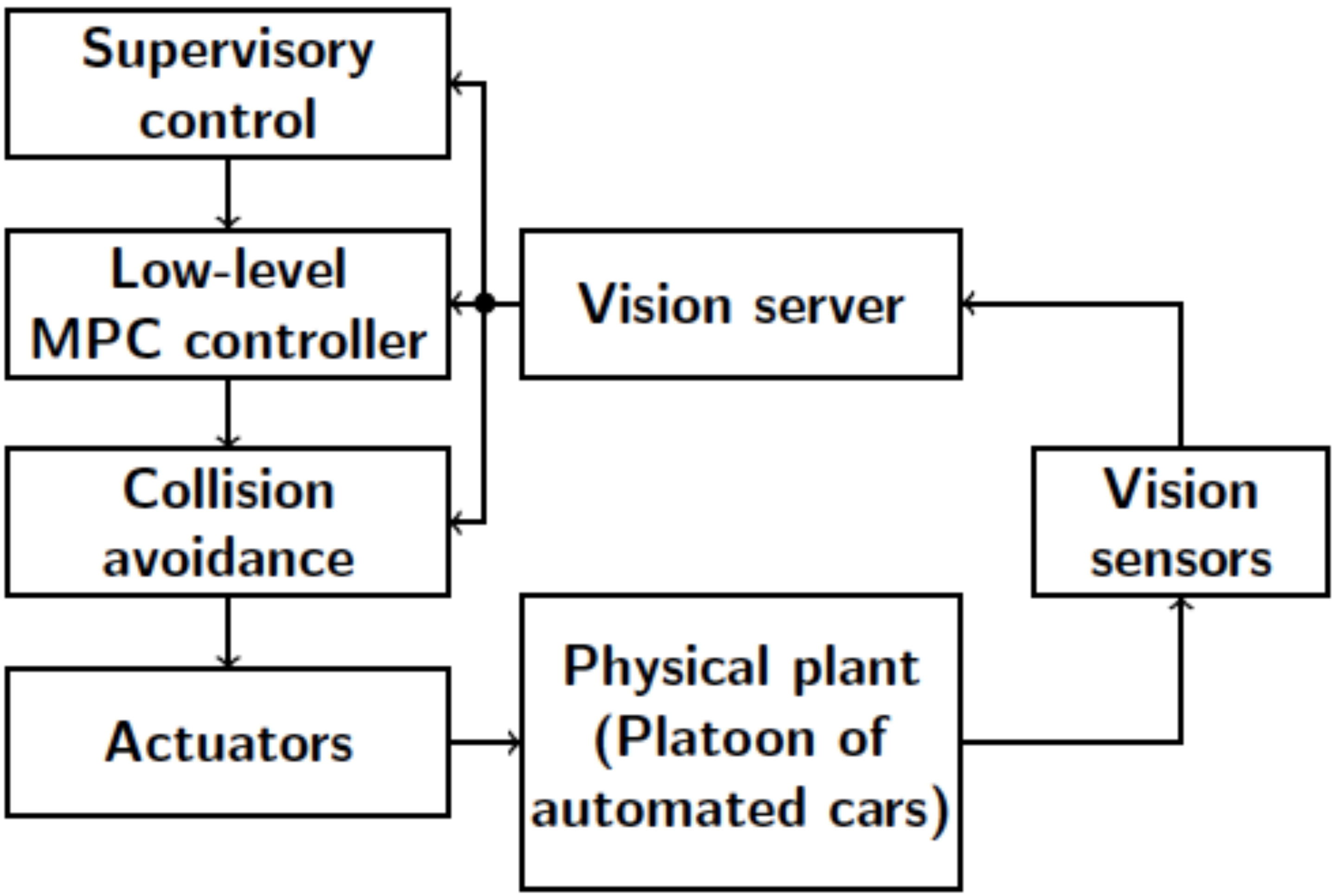
- Car hacking is the future and sooner or later you'll be hit
  - https://www.theguardian.com/technology/2016/aug/28/car-hacking-future-self-driving-security

- Critical reasons for crashes investigated in the national motor vehicle crash causation survey
  - https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115

- Hackers Remotely Kill a Jeep On the Highway- With Me in it”
  - https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

- Feature of daily news …

Testbed architecture



System model for automatic vehicles

- Plant model for vehicle i given by its kinematic equations

$$x_i[t+1] = x_i[t] + h\cos(\theta_i[t])v_i[t] + h\cos(\theta_i[t])w_{ix}[t]$$
$$y_i[t+1] = y_i[t] + h\sin(\theta_i[t])v_i[t] + h\sin(\theta_i[t])w_{iy}[t]$$
$$\theta_i[t+1] = \theta_i[t] + h\omega_i[t] + hw_{i\theta}[t]$$

- h is the sampling period (100ms)
- $v_i[t]$ a control input, denoting speed
- $\omega_i[t]$ a control input, denoting angular
- $w_{ix}[t]$, $\underline{w_{iy}}[t]$, $w_{i\theta}[t]$ all $N(0,2)$, i.i.d.

- Non-linear system

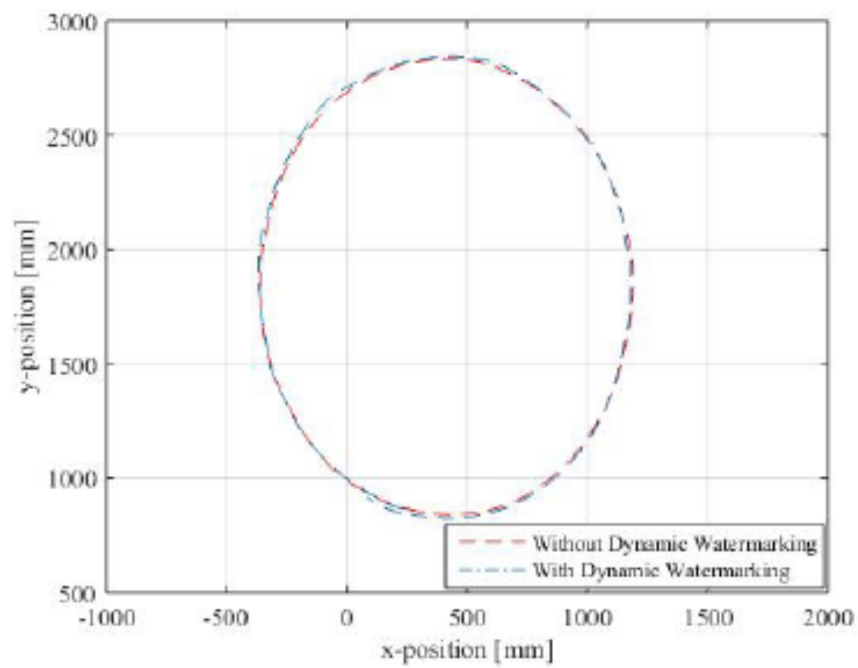
Watermarked system's performance in absence of attack

- Watermarked system

$$x_i[t+1] = x_i[t] + h\cos(\theta_i[t])u_i^g(\mathbf{z}_1^t,\mathbf{z}_2^t) + h\cos(\theta_i[t])e_{iv}[t] + h\cos(\theta_i[t])w_{ix}[t]$$
$$y_i[t+1] = y_i[t] + h\sin(\theta_i[t])u_i^g(\mathbf{z}_1^t,\mathbf{z}_2^t) + h\sin(\theta_i[t])e_{iv}[t] + h\sin(\theta_i[t])w_{iy}[t]$$
$$\theta_i[t+1] = \theta_i[t] + h\omega_i[t] + he_{i\theta}[t] + hw_{i\theta}[t]$$

- Performance with and without watermarking

- Watermarks do not result in any added penalty on performance



Sensor attack

- Sensor attack

$$z_{2x}[t_A] = x_2[t_A] + \tau, \text{ where } \tau = \text{bias}$$

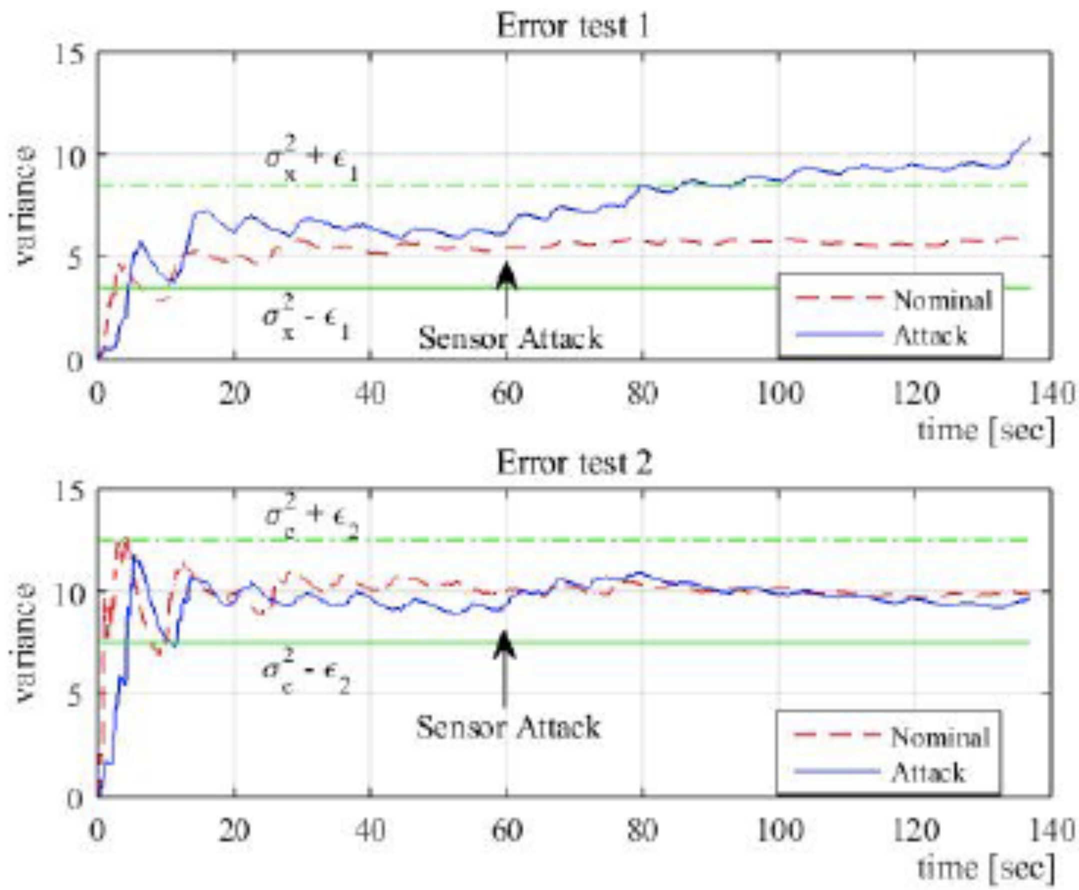
$$z_{2x}[t+1] = z_{2x}[t] + h\cos(\theta_2[t])u_2^g(\mathbf{z}_1^t, \mathbf{z}_2^t) + \cos(\theta_2[t])n[t]$$

$$n[t] \sim \mathcal{N}(0, \sigma_x^2)$$

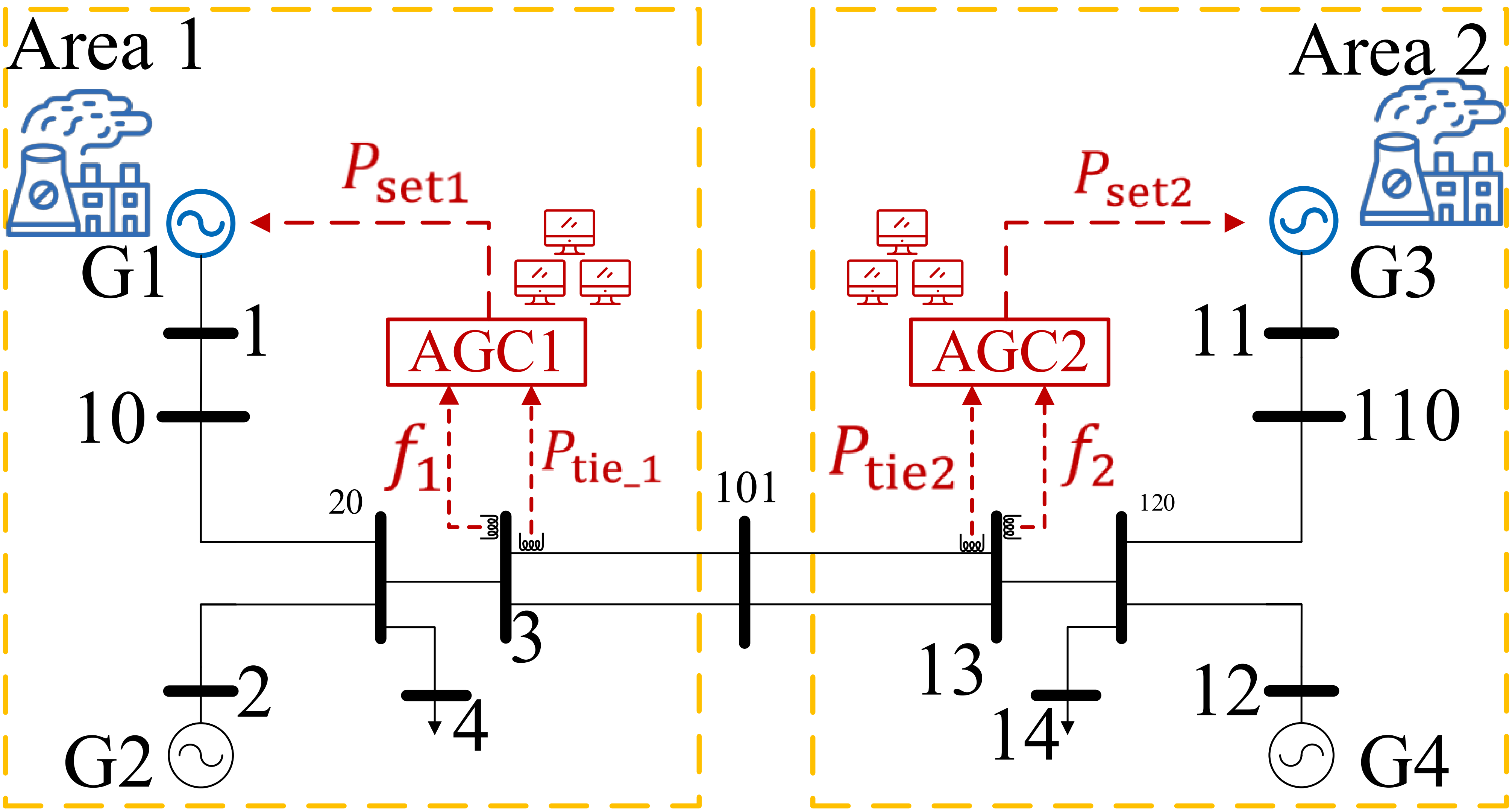
- This attack passes Test 2, but fails Test 1

Test Statistics

- Fails Test 1
- Passes Test 2

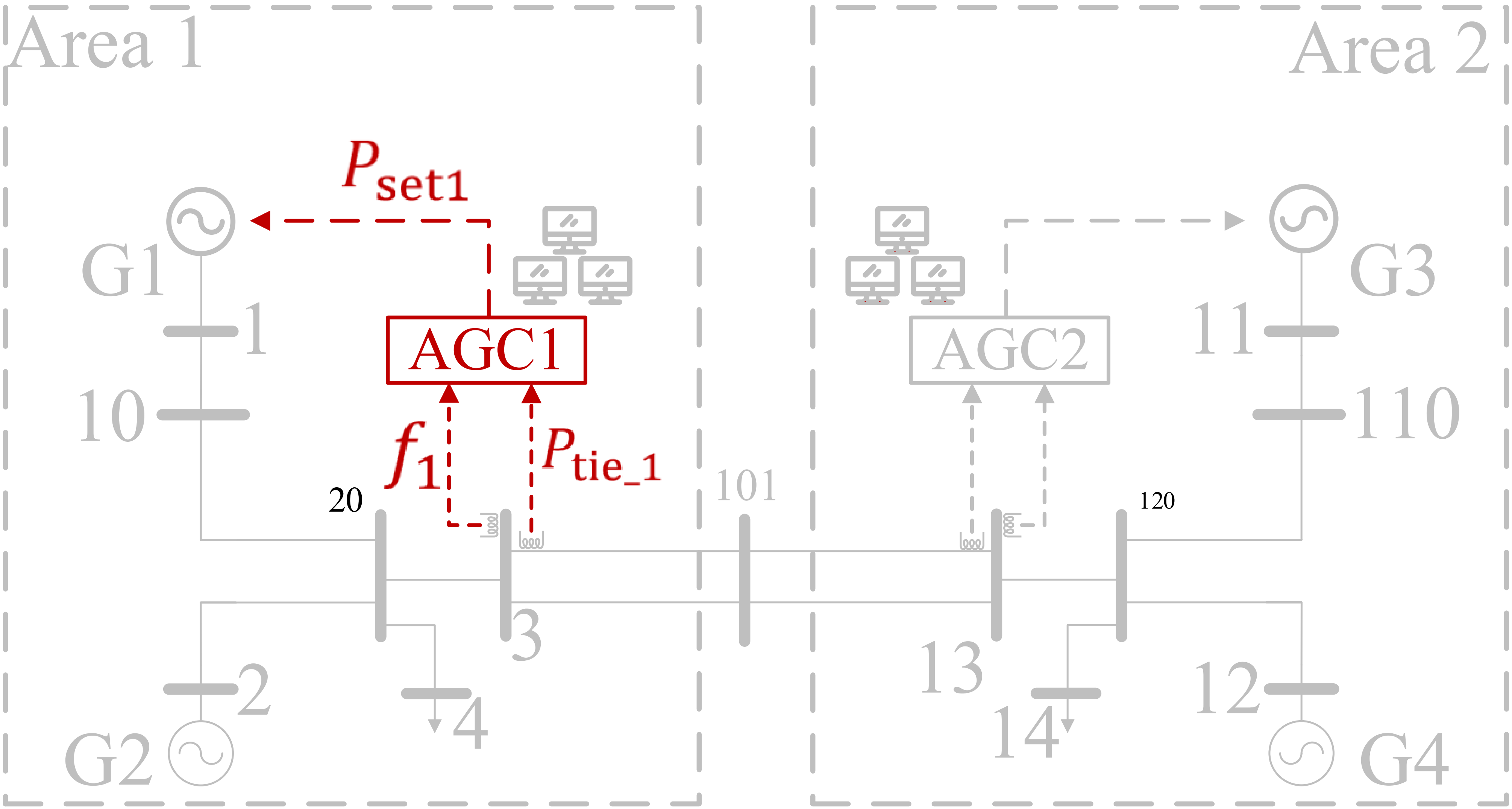


Automatic Generation Control (AGC)

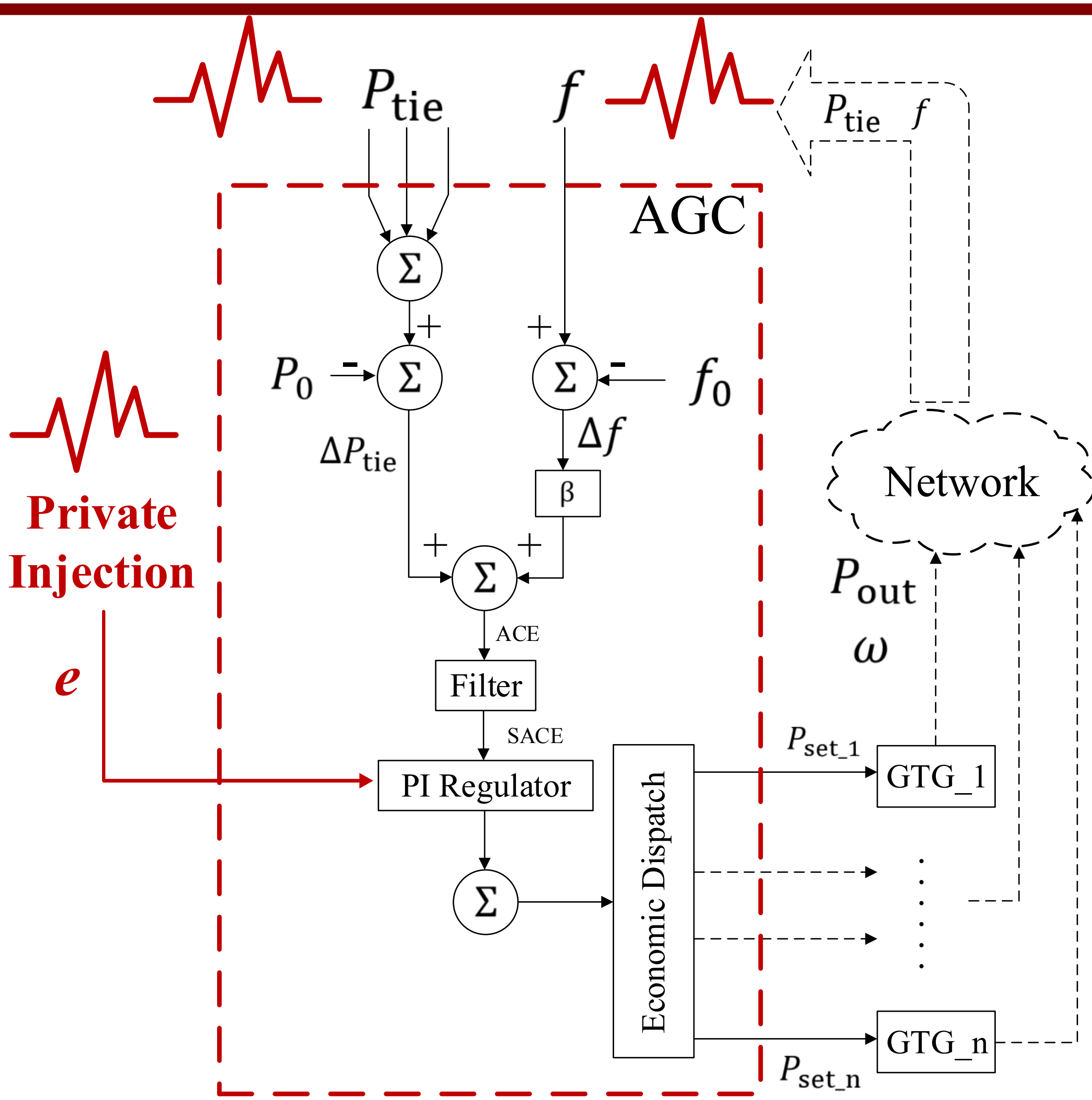


Instead of honestly reporting the real measurement $\boldsymbol{y_i}$, the sensors might be manipulated to report $\boldsymbol{z_i}$, where $\boldsymbol{z_i} \neq \boldsymbol{y_i}$.

Automatic Generation Control (AGC)



Dynamic Watermarking in the Context of AGC



*Certain indelible pattern of **the private injection** is imprinted into the measurement feeding to AGC.*

$\boldsymbol{f}_i(\boldsymbol{y}_i^t)$ $\;f\;$ P_{tie} $\;\boldsymbol{e}[k]$

↓

Kalman Filtering

↓

Two Test in Dynamic Watermarking

↓

Detect attacks based on indicators

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\boldsymbol{\nu}_k\boldsymbol{\nu}_k^T = L_{di}\Sigma_i L_{di}^T$$

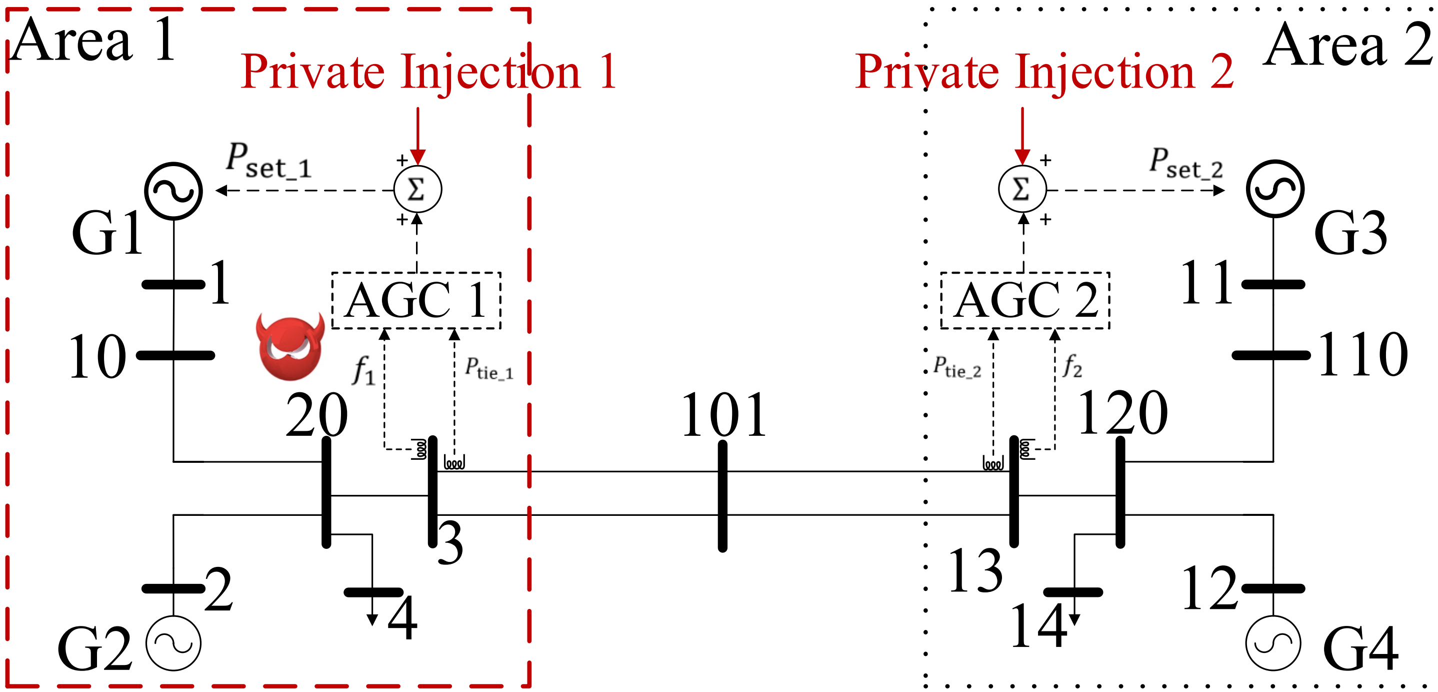
Test 1 $\quad \Sigma_i = C_{di}PC_{di}^T + R.$

Test 2

$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{T}\boldsymbol{e}(k-1)\boldsymbol{\nu}_k^T = 0$$

$$\boldsymbol{\nu}_k = \boldsymbol{x}(k|k) - A_{di}\boldsymbol{x}(k-1|k-1) - B_{di}\boldsymbol{f}_i(\boldsymbol{z}_i^{k-1}) - B_{di}\boldsymbol{e}(k-1)$$

Performance Validation: the Impact of Private Injection



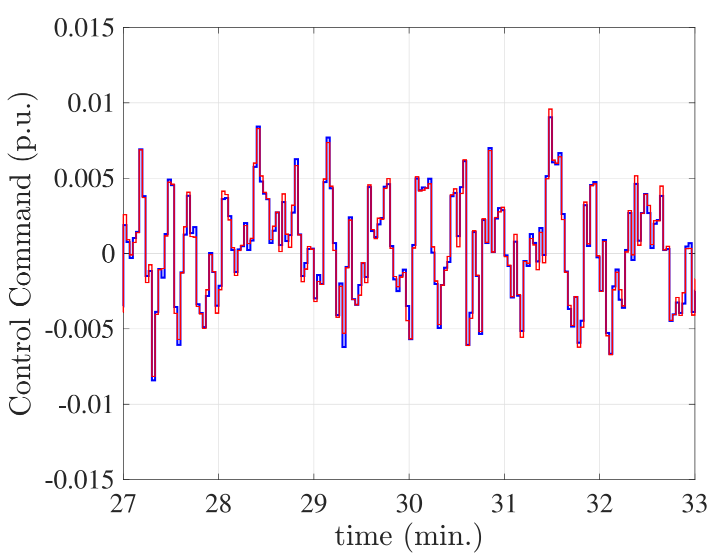
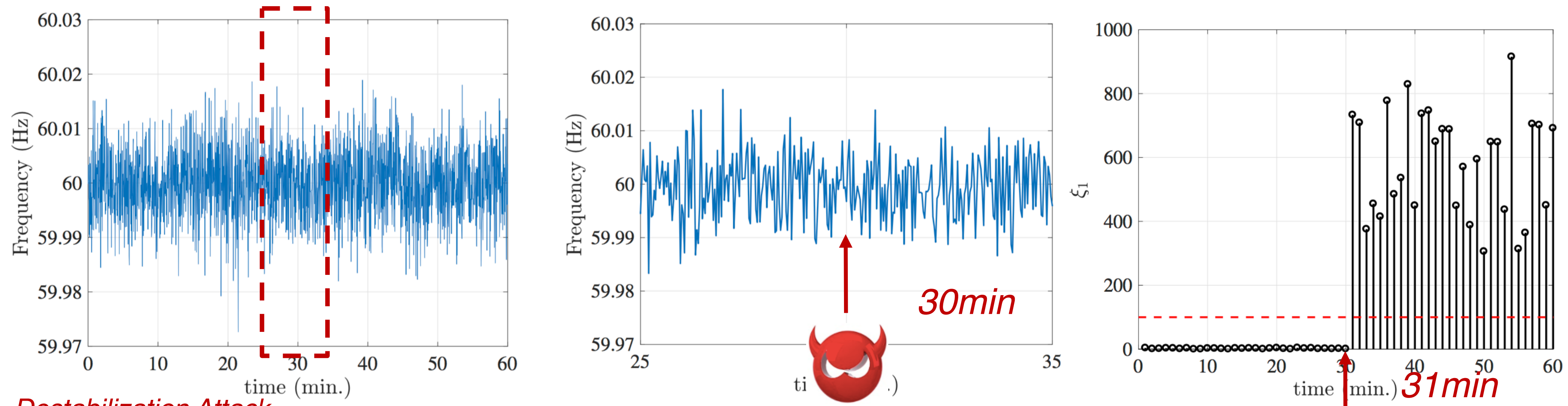


TABLE I
THE IMPACT OF PRIVATE INJECTION ON FREQUENCY AND CONTROL COMMAND OF AGC

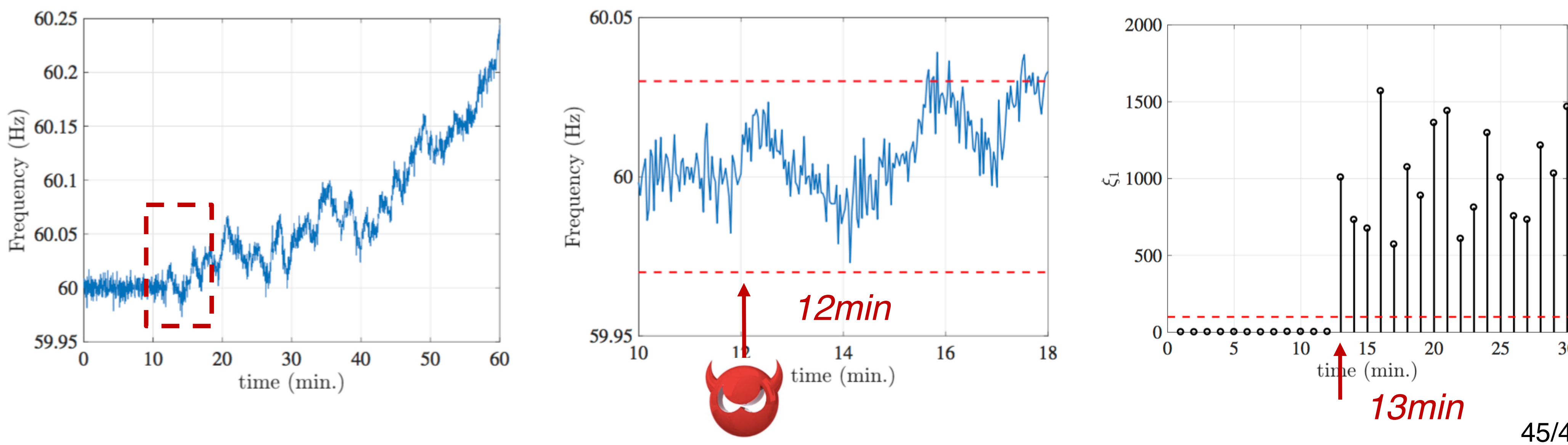
	Control Command	Frequency
Variance without $e(k)$	4.1593×10^{-5}	1.3448×10^{-5}
Variance with $e(k)$	4.1951×10^{-5}	1.3562×10^{-5}
Change of Variance (%)	0.86%	0.85%

Performance Validation under Replay Attack and Destabilization Attack

Replay Attack



Destabilization Attack



Remarks

- CPS is important for society and economy
- Lot of future infrastructure may be CPS
- Societally and economically important
- Security of CPS is a very rapidly emerging area
- Critical for safety of future infrastructure
- Lots of attacks have already been demonstrated

References

- Yilin Mo, and Bruno Sinopoli, "Secure control against replay attacks," Allerton, pp. 911-918. IEEE, 2009.
- Sean Weerakkody, Yilin Mo, and Bruno Sinopoli, "Detecting integrity attacks on control systems using robust physical watermarking," 53rd IEEE Conference on Decision and Control, pp. 3757-3764. IEEE, 2014.
- Yilin Mo, Rohan Chabukswar, and Bruno Sinopoli. "Detecting integrity attacks on SCADA systems," IEEE Transactions on Control Systems Technology, no. 4 (2014): 1396-1407.
- Yilin Mo, Sean Weerakkody, and Bruno Sinopoli, "Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs," IEEE Control Systems 35, no. 1 (2015): 93-109.
- Bharadwaj Satchidanandan and P. R. Kumar, “Dynamic Watermarking: Active Defense of Networked Cyber-Physical Systems.” Proceedings of the IEEE, vol. 105, No. 2, pp. 219-240, February 2017.
- Woo-Hyun Ko, Bharadwaj Satchidanandan and P. R. Kumar, “Theory and Implementation of Dynamic Watermarking for Cybersecurity of Advanced Transportation Systems.” International Workshop on Cyber-Physical Systems Security (CPS-Sec), pp. 235-239, Philadelphia, October 17-19, 2016.
- Bharadwaj Satchidanandan and P. R. Kumar, “Secure Control of Networked Cyber-Physical Systems.” Proceedings of 55th IEEE Conference on Decision and Control, pp. 283-289, December 12–14, 2016, Las Vegas.
- Bharadwaj Satchidanandan and P. R. Kumar, “On Minimal Tests of Sensor Veracity for Dynamic Watermarking-Based Defense of Cyber-Physical Systems.” Proceedings of the 9th International Conference on Communication Systems & Networks (COMSNETS 2017), pp. 23-30, January 4-8, 2017, Bengaluru, India.
- Bharadwaj Satchidanandan and P. R. Kumar, “Defending Cyber-Physical Systems from Sensor At- tacks.” In to appear in From 9th International Conference on Communication Systems & Networks (COMSNETS 2017), Lecture Notes in Computer Science, Springer–Verlag, Berlin.
- Bharadwaj Satchidanandan and P. R. Kumar, “Control Systems Under Attack: The Securable and Unsecurable Subspaces of a Linear Stochastic System.” To appear in Emerging Applications of Control and System Theory, 2017.
- T. Huang, B. Satchidanandan, P. R. Kumar, and L. Xie, "An Online Defense Framework against Cyber Attacks on Automatic Generation Control," Working Paper, TAMU-ECE-2017-02. (Submitted to IEEE Transactions on Power Systems).
- Pasqualetti, Fabio, Florian Dorfler, and Francesco Bullo. ”Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems.” IEEE Control Systems 35, no. 1 (2015): 110-127.
- Pasqualetti, Fabio, Florian Drfler, and Francesco Bullo. ”Attack detection and identification in cyber-physical systems.” IEEE Transactions on Automatic Control 58, no. 11 (2013): 2715- 2729.
- Pasqualetti, Fabio, Florian Drfler, and Francesco Bullo. ”Cyber-physical security via geometric control: Distributed monitoring and malicious attacks.” In Decision and Control (CDC), 2012 IEEE 51st Annual Conference on, pp. 3418-3425. 2012.
- Teixeira, A, Iman Shames, Henrik Sandberg, and Karl H. Johansson. ”Revealing stealthy attacks in control systems.” In Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on, pp. 1806-1813. 2012.

Thank you