A Personal Take on Cyber-Physical Systems

Mo, Yilin (Department of Automation, Tsinghua University)

TC on Cyber-Physical Systems Meetings (Aug 21, 2020)

Introduction

Definition of Cyber-Physical System

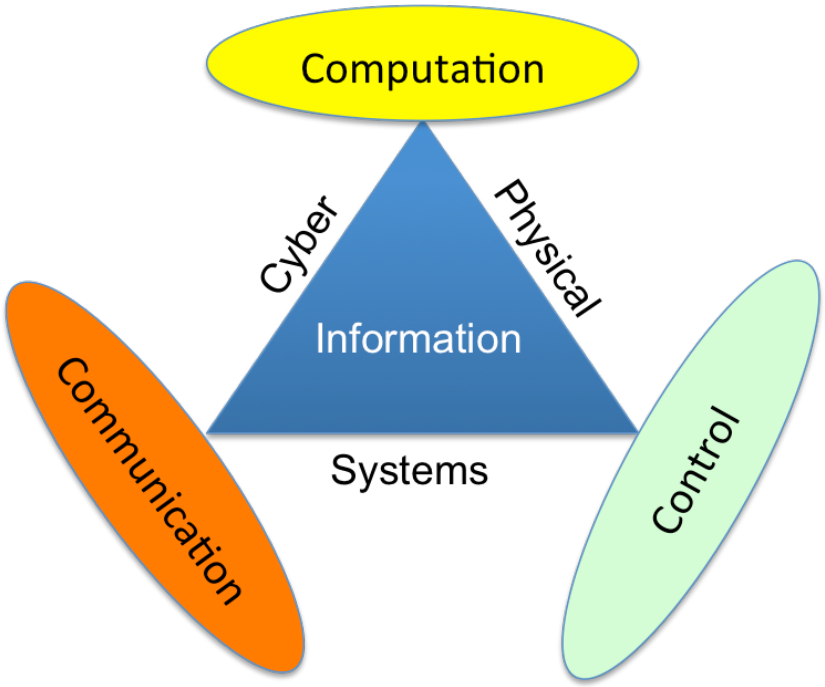


Figure 1: (Figure)

CPSs refer to the next generation of engineered systems that require tight integration of ***computing***, ***communication***, *and* ***control*** *technologies to achieve* ***stability***, ***performance***, ***reliability***, ***robustness***, *and* ***efficiency*** *in dealing with physical systems of many application domains.* [**Kumar**], $Proceedings of the IEEE$, 2012

Applications

Presentation contains image grid. LaTeX export not supported.

Cyber-Physical-Human Energy System

Complex interaction between humans, HVAC and grid. [**Guan2010**], [**Jia2018**]
Presentation contains image grid. LaTeX export not supported.

Convergence of Computation and Control

- Cyber Space

Presentation contains image grid. LaTeX export not supported.

- Physical Space

Warning! Figure omitted as gif format **not** supported in LaTeX: "Figure" (See HTML presentation instead.)

$$\dot{x} = f(x)$$

- 1948: The term "Cybernetics" was coined by Norbert Wiener.

Figure 2: Figure

- 1970s: Real time system
- 1990s: Hybrid system
- 2006: The term “Cyber-Physical System” was coined by NSF.

Convergence of Communication and Control

- 1971: ALOHAnet, first public wireless packet data network
- 1997: IEEE 802.11 (Wifi) was introduced
- 1997: Smart Dust Project proposed by researchers from UCB

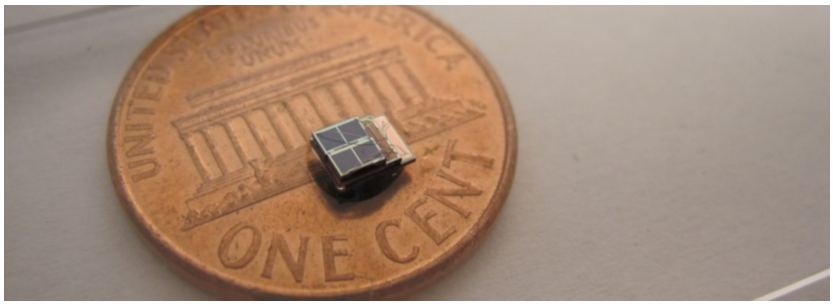
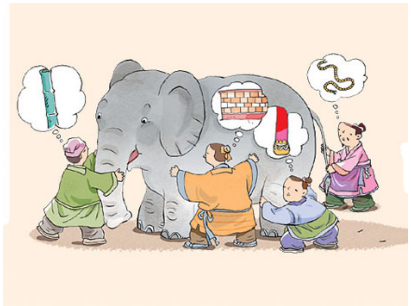


Figure 3: (Figure)

- 1999: The term Internet of Things was invented.
- 2000: Panel on Future Directions in Control, Dynamics, and Systems [**Murray2003**]

Networks of sensory or actuator nodes with computational capabilities, connected wirelessly or by wires, can form an orchestra which controls our physical environment.

Many-Sidedness of CPS



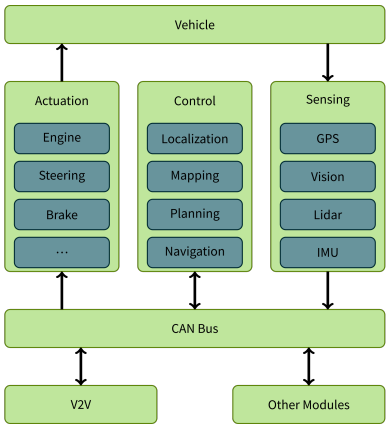


Figure 4: Simplified Diagram of Autonomous Car (Figure)

Networked Control System

Definition of Networked Control System

Networked control systems are spatially distributed systems in which the communication between sensors, actuators, and controllers occurs through a shared band-limited digital communication network. [**Zhang2020**], *Journal of Automatica Sinica*

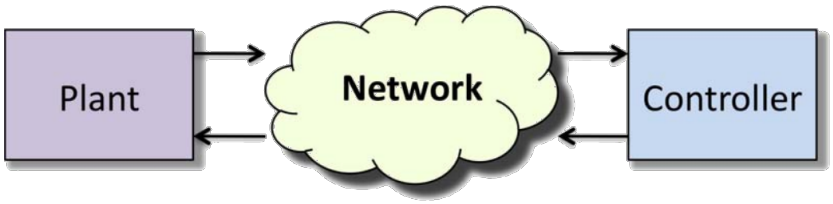


Figure 5: Diagram of Networked Control System (Figure)

- Control over Networks:
  - Communication is **imperfect**: quantization, delay, packet drops
  - Communication is **expensive**: offline scheduling, event-based scheduling
  - **Multiple participants** and **local information**: multi-agent system
- Control of Networks

Estimation and Control over Lossy Channels

Estimation over Lossy Channels

- Problem proposed by Sinopoli et al. [**Sinopoli2004**]
- System model:

$$x_{t+1} = Ax_t + w_t,\ y_t = Cx_t + v_t$$

- Erasure channel model:

$$\tilde{y}_t = \gamma_t y_t$$

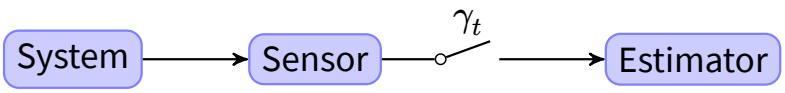


Figure 6: (Figure)

- Goal: given $\gamma_{0:t}$, $\tilde{y}_{0:t}$, estimate x_t.
- Optimal estimator is a Kalman filter with a time varying gain depending on the packet loss process $\{\gamma_t\}$.

- The estimator is stable (in the mean square sense) only when the packet arrival rate is larger than a critical value γ_c.
- A lower bound of λ_c:

$$\lambda_c \geq 1 - \frac{1}{\rho(A)^2}$$

Lower Bound is not Tight

- Counterexample:

$$x_{t+1} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} x_t + w_t,\ y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + v_t.$$

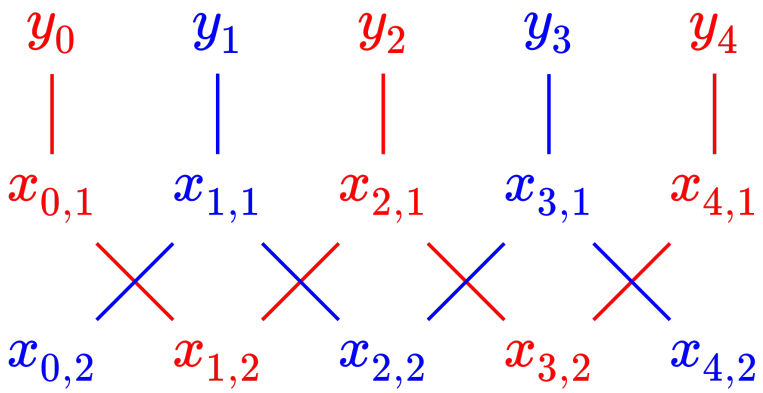


Figure 7: (Figure)

- Critical packet arrival probability is 15/16, not 3/4.
- The lower bound is tight if the system is non-degenerate, i.e., observability is preserved under different sampling frequency [**Mo2012**]
- For degenerate system, the exact critical value is derived in [**Sui2015**]

Control over Lossy Channels [Schenato2007]

- System model:

$$x_{t+1} = Ax_t + B\tilde{u}_t + w_t,\ y_t = Cx_t + v_t$$

- Channel model:

$$\tilde{u}_t = \nu_t u_t, \tilde{y}_t = \gamma_t y_t.$$

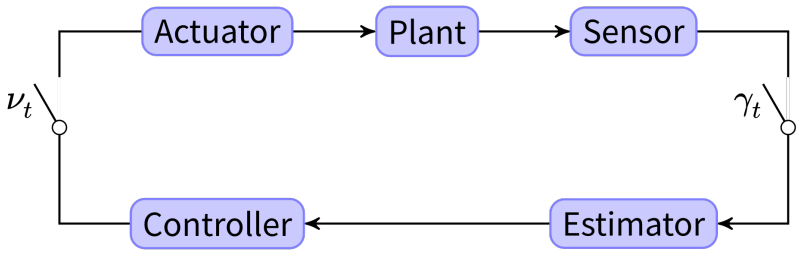


Figure 8: (Figure)

Control over Lossy Channels: TCP case

- TCP information set: $\nu_{0:t}, \gamma_{0:t}, \tilde{y}_{0:t}$
- Optimal control policy (in the LQG sense):
  - Kalman filter with time varying gain
  - Fixed gain state feedback (Gain is derived from an Modified Algebraic Riccati Equation)
  - Separation principle holds
- Stability criteria:
  - Measurement packet arrival rate greater than λ_c
  - MARE has a fixed point solution

Control over Lossy Channels: UDP case

- UDP information set: $\gamma_{0:t}, \tilde{y}_{0:t}$
- Optimal control policy (in the LQG sense):
  - State follows a Gaussian mixture distribution [**Lin2016**]
  - Optimal control is an **open problem**
  - Separation principle does NOT hold
- Stability criteria: **Open problem**

Scheduling for State Estimation

- System model:

$$x_{t+1} = Ax_t + w_t,\ y_t = Cx_t + v_t$$

- Erasure channel model:

$$\tilde{y}_t = \gamma_t y_t$$

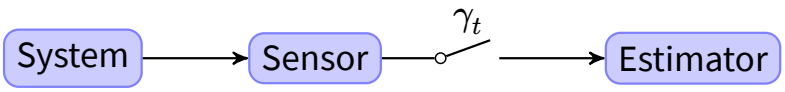


Figure 9: (Figure)

- Goal: Design $\{\gamma_t\}$ to minimize estimation error, while satisfying energy, bandwidth, topological constraints.

Off-line Schedule

- The schedule is based on the statistics of the system, and hence can be determined off-line:
  - Deterministic Schedule: send the measurement only at even time.
  - Stochastic Schedule: send the measurement with 50% probability at each time
- The "optimal" schedule is periodic [**Mo2014**], [**Zhao2014**]
- For special cases, closed form solution can be obtained. [**Shi2011**]
- For general cases, convex relaxation [**Mo2011**] or submodularity [**Shamaiah2010**] can be exploited.

Event-based Schedule

- The schedule depends on both the statistics and the realization of the

system.

- For example: send the temperature if the temperature is outside $[25, 30]$.
- *No news is good news:* If receiving no measurement at time t, the fusion center knows the measurement is within $[25, 30]$

Deterministic Trigger

Presentation contains image grid. LaTeX export not supported.

- When $\gamma_t = 0$, y_t follows a truncated Gaussian distribution.
- The optimal filter is given by the forward algorithm of hidden Markov model.
- Need to keep track of the pdf of x_t given $y_{0:t}$.
- Near optimal filter with low complexity can be used. [**Wu2013**], [**Shi2014**]

Stochastic Trigger

Presentation contains image grid. LaTeX export not supported.

- At each time k, the sensor generates a random variable $\zeta_t \sim U[0,1]$
- Decide whether to send or not based on the following rule:

$$\gamma_t = \begin{cases} 0 & \zeta_t \leq \Phi(y_t) \\ 1 & \zeta_t > \Phi(y_t) \end{cases}, \text{ with } \Phi(y) = \exp\left(-\frac{1}{2} y' Y y\right).$$

- The closed-form solution of the optimal filter is a Kalman like filter [**Han2013**]

CPS Security and Privacy

Security Risks

- The next generation CPS: **Smart Grids**, **Smart Buildings**, **Internet of Things**, will make extensive use of widespread sensing and networking.
- As the CPSs become "*smarter*", they are also more vulnerable to malicious attacks.

Presentation contains image grid. LaTeX export not supported.

The First CPS Malware: Stuxnet

- Stuxnet is the first discovered malware that spies on and subverts industrial control systems. It was discovered in June 2010.

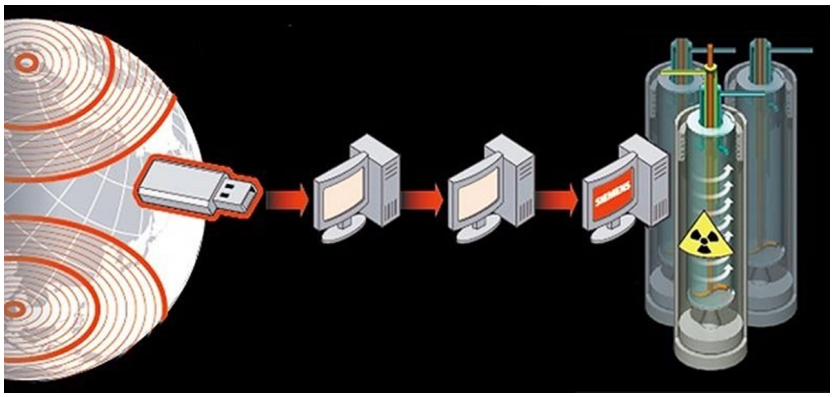


Figure 10: Stuxnet Incident (Figure)

Threats to Industrial Control Systems

- In FY 2016, ICS-CERT received **290** incidents. The scope of incidents includes:

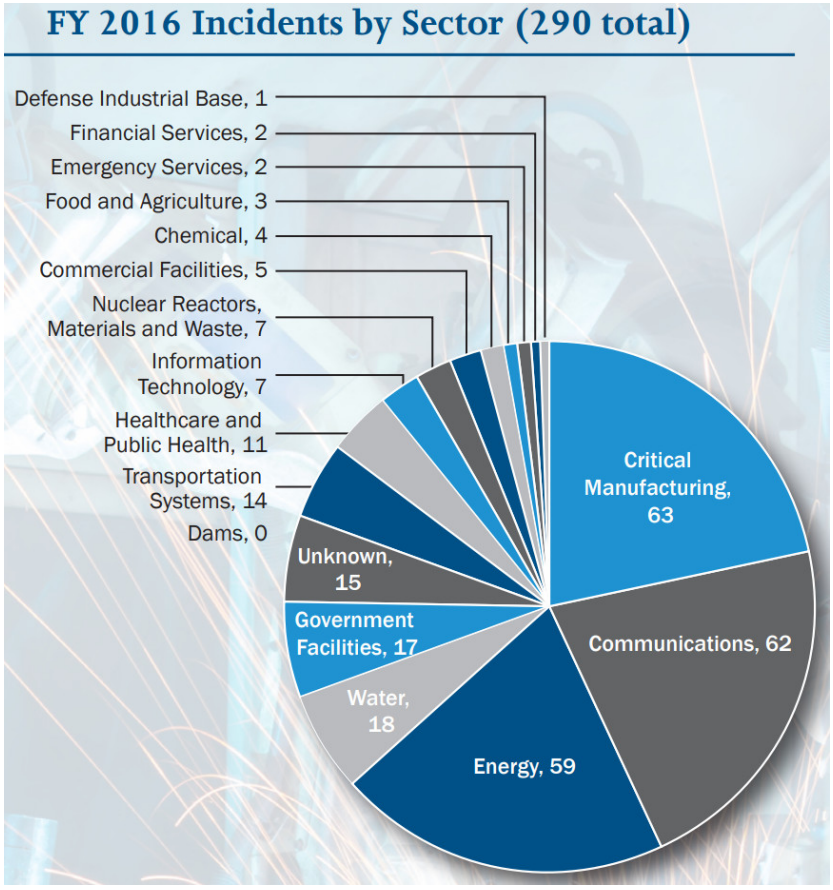


Figure 11: Figure

- Unauthorized access,
- Exploitation of zero-day vulnerabilities,
- Malware infections within air-gapped control system networks.

Attack Through Supply Chain

- 70% of the components for Boeing 787 are manufactured by other suppliers [**Tang2009**]

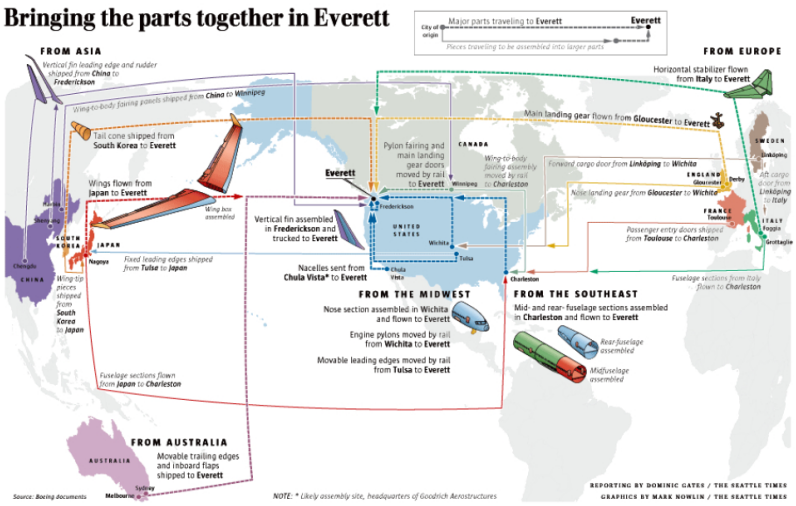


Figure 12: Supply Chain for Boeing 787 (Figure)

Black Energy

- A Successful attack on **critical infrastructure** can be devastating.

Presentation contains image grid. LaTeX export not supported.

Cyber-Physical Security

- CIA Triad: **Confidentiality**, **Integrity**, **Availability**
- A case study for smart grid [**Mo2011**]

	Control Signal	Measurements	Software
Confidentiality	Exposure of System Structure	Unauthorized Access	Piracy
Integrity	**Changes of Control Command**	**Incorrect Data**	**Malicious S**
Availability	**Inability to Control the Grid**	**Unavailability of Measurements**	N/A

Threat Modelling

- [**Teixeira2015**] proposes a 3-dimension attack space:

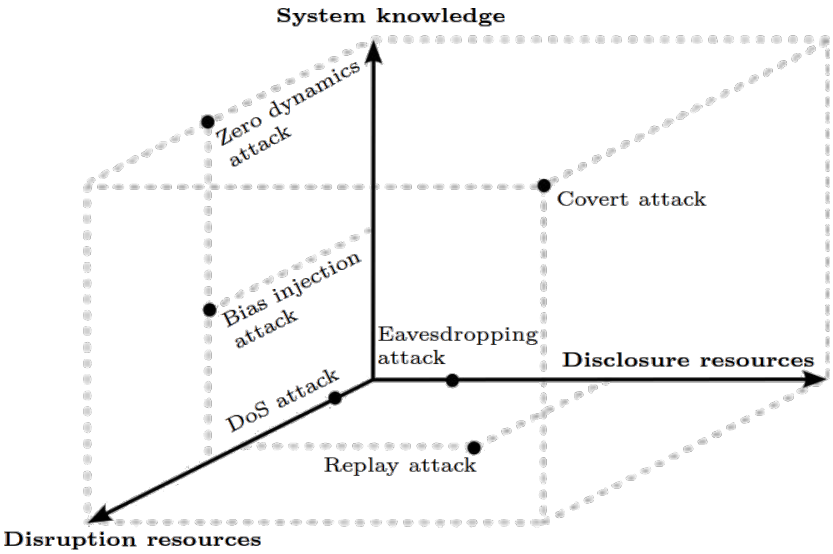


Figure 13: Figure

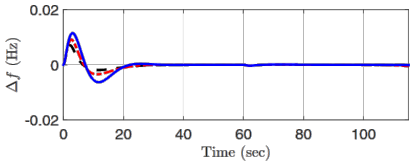
  - **System Knowledge**: Knowledge of off-line system parameters
  - **Disclosure Resources**: Knowledge of on-line signals
  - **Disruption Resources**: Manipulation of on-line signals

- Can disclosure resources be used to gain system knowledge (data-driven attack)? [**Park2019**], [**Yuan2020**]

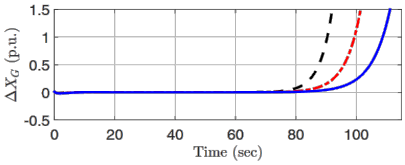
Threat Analysis and Design

Adversary's Perspective:

- Maximize impact
- Minimize required resources
- Stealthiness constraints



(a) Incremental frequency deviation Δf (Hz)



(b) Change in valve position ΔX_G (p.u.)

Figure 14: Figure

System's Perspective:

- Fundamental limit
- Stability under attack
- Performance under attack
- Cost of Security

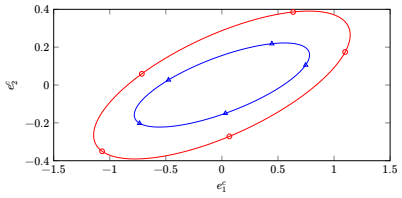


Figure 15: Figure

Countermeasures: Defense in Depth

- Defense in Depth

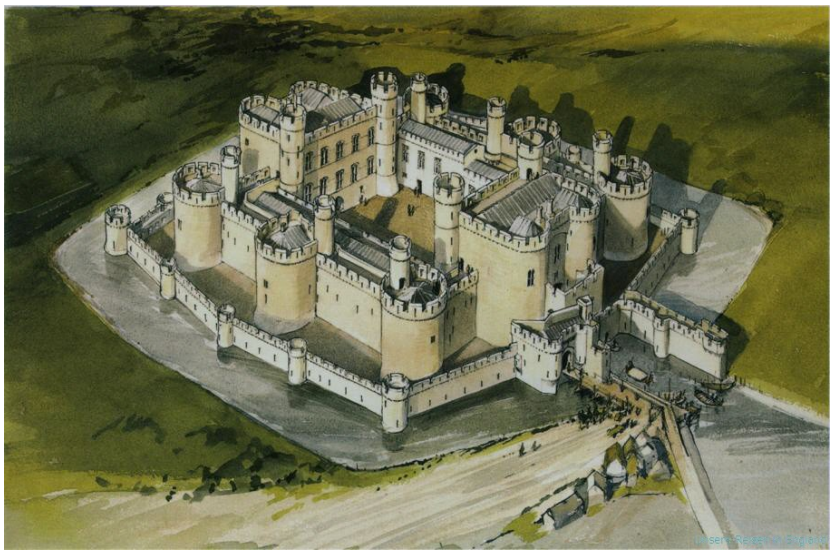


Figure 16: Figure

- Prevention
  - Using cryptography/coding to preserve confidentiality (privacy)
- Detection
  - Fault Detection and Isolation (FDI)
  - Active Detection
- Resiliency
  - Off-line: Resilient system design
  - On-line: Secure information fusion, control
- Recovery
  - Software rejuvenation
  - Patching
- ...

Active Detection against Replay Attack

Stuxnet

- NY times: *The worm itself now appears to have included two major components. One was designed to send Iran's nuclear centrifuges spinning wildly out of control. Another seems right out of the movies:* ***The computer program also secretly recorded what normal operations at the nuclear plant looked like, then played those readings back to plant operators****, like a pre-recorded security tape in a bank heist, so that it would appear that everything was operating normally while the centrifuges were actually tearing themselves apart.*

System Description

[width=.9]./replaydiagramone

[width=.9]./replaydiagramtwo

- Assumptions: Linear Gaussian systems, Linear state estimator+state feedback, χ_2 failure detector.

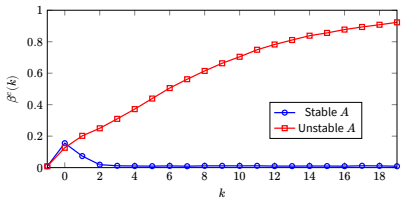


Figure 17: Detection Rate v.s. Time (Figure)

Not all system can detect replay attack!

Active Detection via Physical Watermarking

- Change the control law by adding a zero mean i.i.d. Gaussian watermarking signal $\zeta(k)$:

$$u(k) = \text{Optimal LQG Control} + \zeta(k).$$

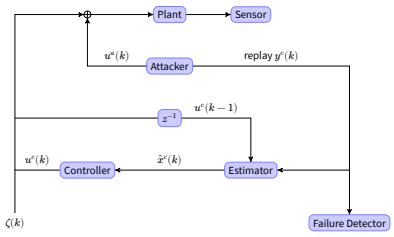


Figure 18: Watermark (Figure)

- Optimizing the covariance of $\zeta(k)$:

$$\begin{aligned} &max \quad Detection\ Performance \\ &\text{s.t.} \quad Ctrl\ Loss \leq \delta \end{aligned}$$

- Can be relaxed into semidefinite programming [**Mo2014**]
- The problem can be solved online, without explicit system knowledge [**Liu2020**]
- Other watermarking schemes (e.g., multiplicative watermark) [**Ferrari2020**], [**Satchidanandan2017**]

Simulation: Tennessee Eastman Process

- The TEP is a realistic industrial model for process control.

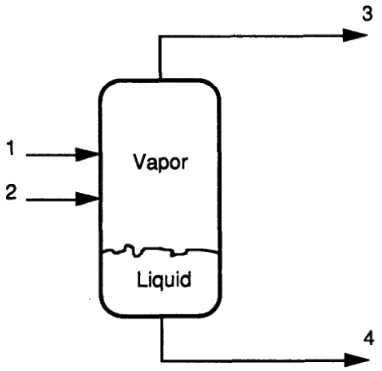


Figure 19: TEP Model (Figure)

- The simplified model contain 4 inputs, 4 outputs and 7 internal states.

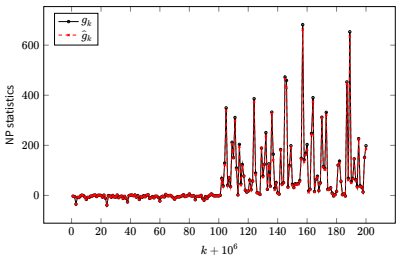


Figure 20: Detection Performance (Figure)

- Physical watermarks enable replay attack detection.
- The optimal watermark signal and detector can be learned using data-driven approach.

Secure State Estimation

Static State Estimation

- Sensor model:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = Hx + noise + attack.$$

- The noise is **small** (zero mean Gaussian or bounded) but **ubiquitous**.
- The attack is p-**sparse** but can be **arbitrarily large**.
- Originally proposed by Peng et al.[**Liu2011**] in the context of DC power flow model.

Fundamental Limit for Noiseless Case

- The system is called p-observable if H is full column rank even after removing rows corresponding to an arbitrary set of p sensors.
- The system is NOT p-observable $\Rightarrow$ There exists undetectable attack.
- The system is NOT $2p$-observable $\Rightarrow$ There exists unidentifiable attack.
- The system is NOT $2p$-observable $\Rightarrow$ One cannot "securely" estimate the state with bounded error.

- The "secure" estimation problem is **NP-hard** in general [**Hendrickx2014**], [**Mao2019**]

Secure Static Estimator

- Estimator with Combinatorial Complexity [**Fawzi2014**], [**Ren2020**], e.g.,

$$\begin{aligned} \underset{\hat{x},a,w}{minimize} \quad & \|w\|^2 \\ \text{subject to} \quad & y = H\hat{x} + w + a, \\ & \|a\|_0 \leq p. \end{aligned}$$

- Require $2p$-observable to be "secure"
- Achieves fundamental limit

Simulation: IEEE 14-Bus System

- The system has 27 sensors and 13 states

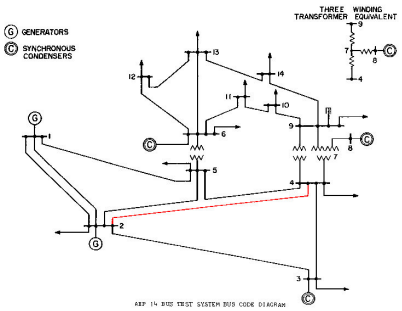


Figure 21: IEEE 14-Bus System (Figure)

- The power flow meter on the red line is being attacked

- Convex optimization based estimator [**Han2019**]

$$\hat{x} = \underset{\hat{x}}{argmin}\, f_i(y_i - H_i\hat{x})$$

- Require stronger condition than $2p$-observable to generate "secure" estimate.

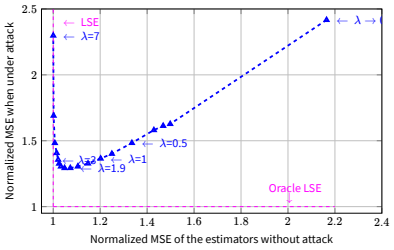


Figure 22: Stuxnet Incident (Figure)

- With fine-tuned parameters, we can design a "good" estimator both in the **absence** and in the **presence** of attacks.

Dynamic State Estimation

- System Model:

$$x_{t+1} = Ax_t + w_t,\ y_t = Cx_t + v_t + a_t$$

- p-sparse attack model: At most p sensors are compromised ($a_i(t) \neq 0$) during operation.
- Fundamental limit: The state can be "securely" estimated with bounded error only if the system is $2p$-detectable [**Nakahira2018**]

Dynamic Estimator: Moving Horizon Approach

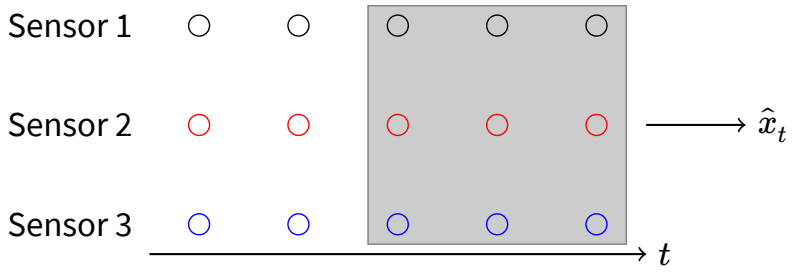


Figure 23: Moving Horizon with Window Size 3 (Figure)

[**Fawzi2014**], [**Shoukry2017**]

Dynamic Estimator: Local Fusion Approach

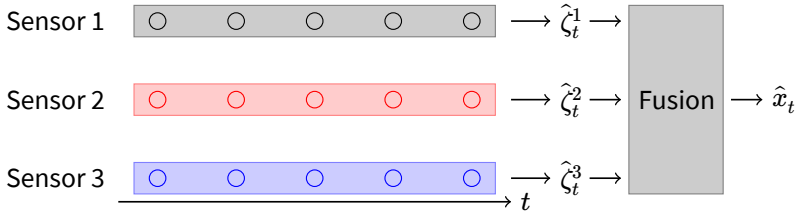


Figure 24: Secure Estimator Design with Local Estimators (Figure)

[**Liu2017**], [**Mao2019**]

Dynamic Estimator: Switching Approach

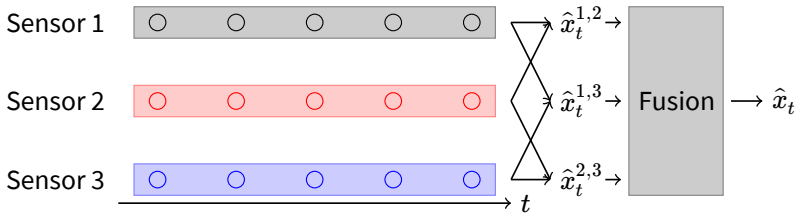


Figure 25: Secure Estimator Design with Switching (Figure)

[**Nakahira2018**], [**An2018**]

Case Study: Secure Est against GPS Spoofing for Vehicle Localization

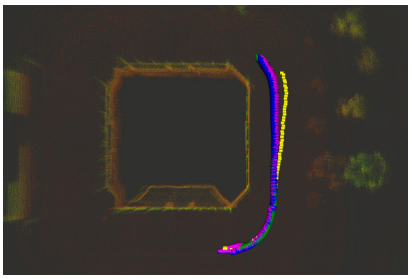


Figure 26: EKF under GPS Spoofing (Figure)

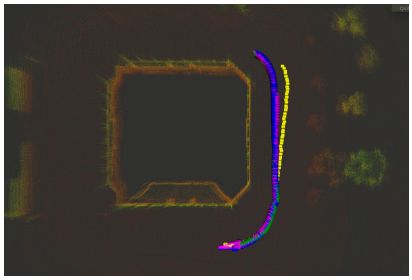


Figure 27: Secure Filter under GPS Spoofing (Figure)

Privacy

Differential Privacy

- ϵ-private: Let $\mathcal{M}$ be a randomized algorithm that takes certain data x as input. $\mathcal{M}$ is ϵ-differentially private if

$$\exp(-\epsilon)P(\mathcal{A}(x_2) \in S) \leq P(\mathcal{A}(x_1) \in S) \leq \exp(\epsilon)P(\mathcal{A}(x_2) \in S),$$

for any adjacent x_1 and x_2 and measurable set S.

- Other privacy metrics exists, e.g., (ϵ, δ)-privacy, mutual information, information leakage, ...
- Trade-off: Utility v.s. Privacy

Additive Noise Mechanism

- Laplacian Mechanism:

$$\mathcal{M}_{f,\epsilon}(x) = f(x) + Lap$$

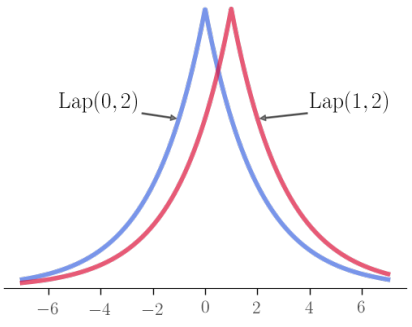


Figure 28: (Figure)

- Gaussian Mechanism:

$$\mathcal{M}_{f,\epsilon}(x) = f(x) + \mathcal{N}$$

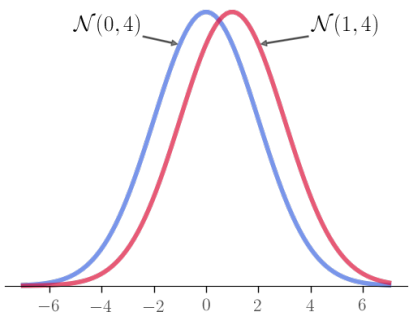


Figure 29: (Figure)

Privacy Preserving Consensus

- The goal:

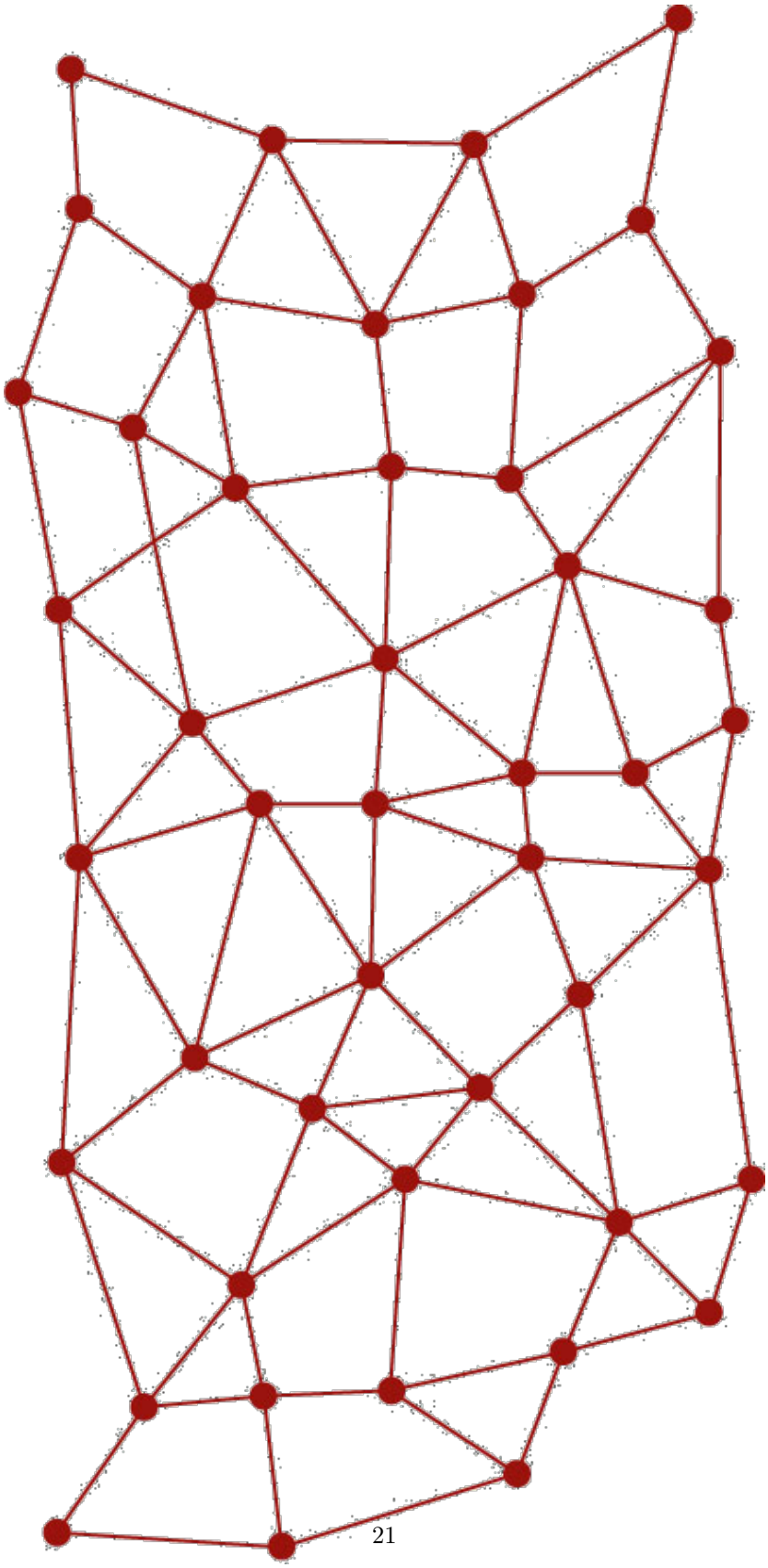


Figure 30: Figure

  - Utility: Converges to the average of the initial states
  - Privacy: Not revealing the exact initial state to other agents

- Algorithm: [**Mo2017**]

$$x_i^+(k) = x_i(k) + v_i(k),\ x_i(k+1) = a_{ii}x_i^+(k) + \sum_{j\in\mathcal{N}(i)} a_{ij}x_j^+(k)$$

- Utility: Converges to the exact average (as fast as the noiseless consensus)
- Privacy: Privacy of an agent is preserved as long as it has no super neighbor.
- Additive noise mechanism for other distributed algorithms, e.g., estimation, optimization [**Liu2019**], [**Wang2019**], [**He2020**], [**LeNy2014**], [**Cortes2016**]

Homomorphic Encryption

- Partially Homomorphic Encryption:
  - There exists $\oplus$, s.t.: $\mathcal{E}(m_1) \oplus \mathcal{E}(m_2) = \mathcal{E}(m_1 + m_2)$;
  - Or there exists $\otimes$, s.t.: $\mathcal{E}(m_1) \otimes \mathcal{E}(m_2) = \mathcal{E}(m_1 \times m_2)$
  - One additively homomorphic encryption is Paillier cryptosystem.
  - Given A and $\mathcal{E}(x)$, one can compute $\mathcal{E}(Ax)$, if additively homomorphic.
- Fully Homomorphic Encryption: [**Gentry2009**]
  - Both $\oplus, \otimes$ exist
  - Computationally expensive
- Control algorithms based on homomorphic encryption: [**Hadjicostis2019**], [**Shoukry2015**], [**Fang2018**], [**Yan2020**]

Conclusion

Modelling and Design of Cyber-Physical System

All models are wrong, but some are useful. – George Box

- What is a good control model/abstraction for computation and communication?
- Modular versus Cross-Layered Design **Warning!** Figure omitted as gif format **not** supported in LaTeX: "Figure"
  (See HTML presentation instead.)
- Model-based versus Data-driven?

Towards a Science of Cyber-Physical System

- Much important work remains to be done: [**Kumar**]
  - To capture and analyze the dynamics of the communications, computation, control, and applications in a unified theoretical framework.
  - To understand and predict complex behaviors caused by tight interactions between cyber and physical domains.
  - High-level decision making based on information collected from different sources at different spatial and temporal scales

Thank you for your time!

Bibliography