Secure Control Against Replay Attacks

Bruno Sinopoli, Yilin Mo

Department of Electrical and Computer Engineering,
Carnegie Mellon

Trust Autumn 2009 Conference

Today's Infrastructure is...

- unsafe
- aging
- inefficient
- insecure

Monitoring, assessment and control

- Public and private sector enterprises today are highly dependent on information systems to carry out their missions and business functions
- Technological developments have seen these traditionally closed systems become open and internet-connected, thus putting national services critical infrastructure at risk
- Physical infrastructures are poorly monitored and loosely controlled
- Cost of sensing, computing and communication have been major inhibitors
- Management of large scale, distributed spaces is complex
- There is a need for design methodologies to address integrated design of the physical and the cyber system

Cyber Physical System

- Cyber Physical System (CPS) refer to the embedding of widespread sensing, computation, communication and control into physical space.
- *"CPS will transform how we interact with the physical world just like the Internet transformed how we interact with one another."*
- Typical applications of CPS: aerospace, chemical processes, civil infrastructure, energy, manufacturing and transportation.
- Lots of safety-critical applications.

Challenges

- Robust design, verification and analysis of performance
  - Design time
    - Compositional analysis
    - Cross layer design tools
    - Software verification
  - Run time
    - Distributed Event Detection
    - Reconfiguration, adaptability
- Security
  - Attack definition
  - Attack detection
  - Operational continuity
  - Remediation / Reconfiguration
  - Restoration

Information Security

Key properties of information security:

1. Confidentiality: attacker cannot read data packets.
2. Integrity: attacker cannot modify data packets.
3. Availability: data packets are available for estimation and control purpose.
4. ...

Information Security

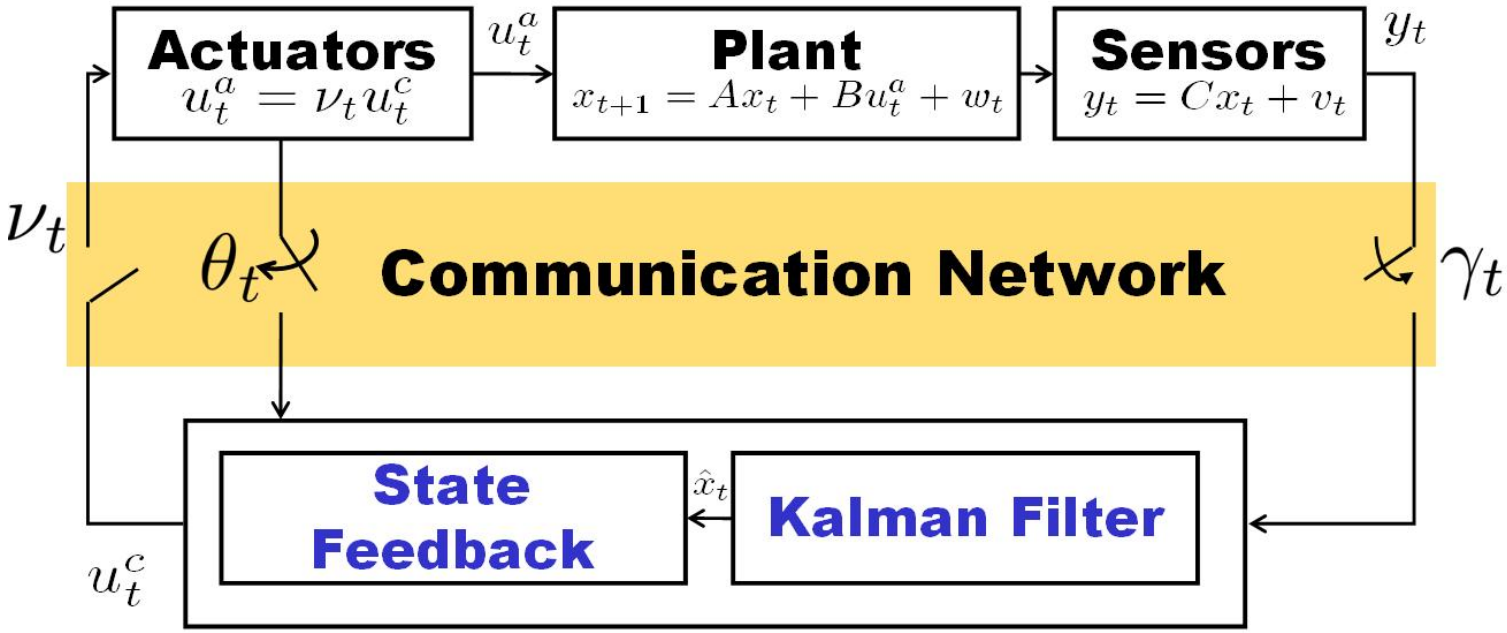
Key properties of information security:

1. Confidentiality: attacker cannot read data packets.
2. Integrity: attacker cannot modify data packets.
3. Availability: data packets are available for estimation and control purpose.
4. ...

Is information security enough?

Attack on Availability

Observation and control packet are sent to the estimator and controller through networks.



Attack on Availability

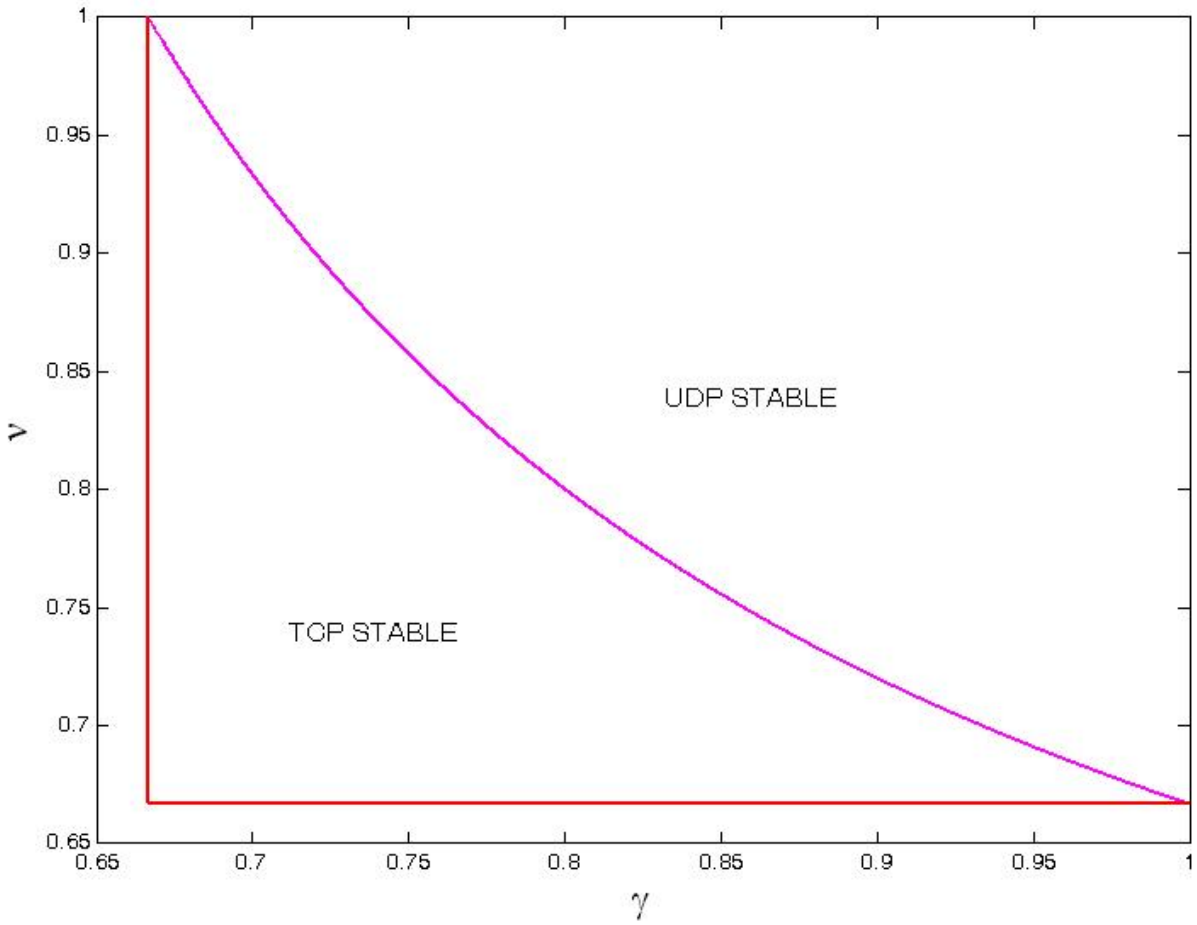


Figure: Stability Regions

Cyber-Physical Attacks

- How to get sensors' readings? Attacker can place a sensor besides CPS sensor.
- How to modify sensors' readings? Attacker can place an actuator besides CPS sensor.
- By monitoring and affecting the environment around CPS sensors, the attack can get and modify the sensors' reading without interfering the communication.

Example: Using thermometer and ice to monitor and control a temperature sensor without breaking confidentiality and integrity.

Attack Model

The attacker can

1. record and modify the sensors' readings y_k
2. inject malicious control input

Attack Model

The attacker can

1. record and modify the sensors' readings y_k
2. inject malicious control input

Replay Attack

1. Record sufficient number of y_ks without adding control inputs.
2. Inject malicious control input and replay the previous y_ks. We denote the replayed measurements as y'_k.

Attack Model

The attacker can

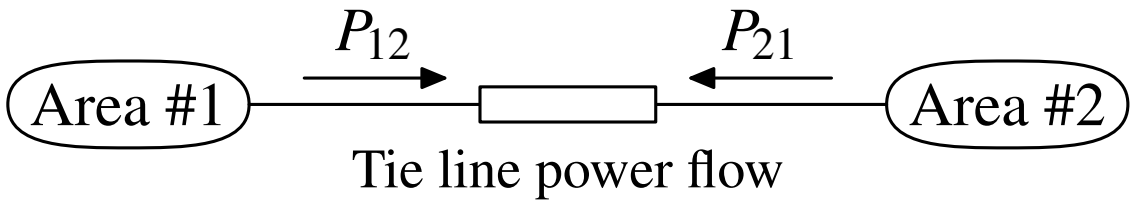
1. record and modify the sensors' readings y_k
2. inject malicious control input

Replay Attack

1. Record sufficient number of y_ks without adding control inputs.
2. Inject malicious control input and replay the previous y_ks. We denote the replayed measurements as y'_k.

When replay begins, there is no information from the systems to the controller. As a result, the controller cannot guarantee any close-loop control performance. The only chance is to detect the replay.

Example: Load Frequency Control in Power System



System Description

$$\frac{d\Delta f_1}{dt} = \alpha_1(\Delta P_{m_1} - \Delta P_{l_1} - P_{12}) + \beta_1 \Delta f_1,$$
$$\frac{d\Delta f_2}{dt} = \alpha_2(\Delta P_{m_2} - \Delta P_{l_2} - P_{21}) + \beta_2 \Delta f_2,$$
$$\frac{d\Delta \varphi}{dt} = 2\pi(\Delta f_1 - \Delta f_2)$$
$$P_{12} = -P_{21} = P_T sin(\Delta \varphi)$$

Example: Load Frequency Control in Power System

- f_i is the frequency in area $\#i$ and Δf_i denotes the deviation from nominal value;
- $\Delta\varphi$ is the difference in phase angle of two areas;
- P_{m_i} and P_{l_i} are the generated and consumed power in area $\#i$, and Δ operator denotes the deviation from nominal value;
- P_{ij} is the power transmitted from area $\#i$ to area $\#j$;
- α_i, β_i are system constants.

1. States (frequency and phase angle): Δf_1, Δf_2, $\Delta\varphi$;
2. Control (generated power): ΔP_{m_1}, ΔP_{m_2};
3. Disturbance (load): ΔP_{l_1}, ΔP_{l_2}.

System Model

We consider the CPS is monitoring the following LTI(Linear Time-Invariant) system

System Description

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k + w_k, \\ y_k &= Cx_k + v_k. \end{aligned} \tag{1}$$

- $x_k \in \mathbb{R}^n$ is the state vector.
- $y_k \in \mathbb{R}^m$ is the measurements from the sensors.
- $u_k \in \mathbb{R}^p$ is the control inputs.
- w_k, v_k, x_0 are independent Gaussian random variables, and $x_0 \sim \mathcal{N}(\bar{x}_0,\ \Sigma)$, $w_k \sim \mathcal{N}(0,\ Q)$ and $v_k \sim \mathcal{N}(0,\ R)$.

Kalman Filter and LQG Controller

- Kalman filter (Assume already in steady states)

$$\hat{x}_{0|-1} = \bar{x}_0, \hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k, \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K(y_{k+1} - C\hat{x}_{k+1|k}).$$

- The LQG controller tries to minimize

$$J = \min \lim_{T\to\infty} E\frac{1}{T}\left[\sum_{k=0}^{T-1}(x_k^T W x_k + u_k^T U u_k)\right].$$

- The solution is a fixed gain controller

$$u_k^* = -(B^T S B + U)^{-1} B^T S A \hat{x}_{k|k} = L\hat{x}_{k|k},$$

where

$$S = A^T S A + W - A^T S B(B^T S B + U)^{-1} B^T S A.$$

χ^2 Failure Detector

The innovation of Kalman filter $y_k - C\hat{x}_{k|k-1}$ is i.i.d. Gaussian distributed with zero mean.

χ^2 Detector

$$g_k = \sum_{i=k-\mathcal{T}+1}^{k} (y_i - C\hat{x}_{i|i-1})^T \mathcal{P}^{-1} (y_i - C\hat{x}_{i|i-1}) \leqslant \textit{threshold},$$

where $\mathcal{P}$ is the covariance of the innovation.

System Diagram

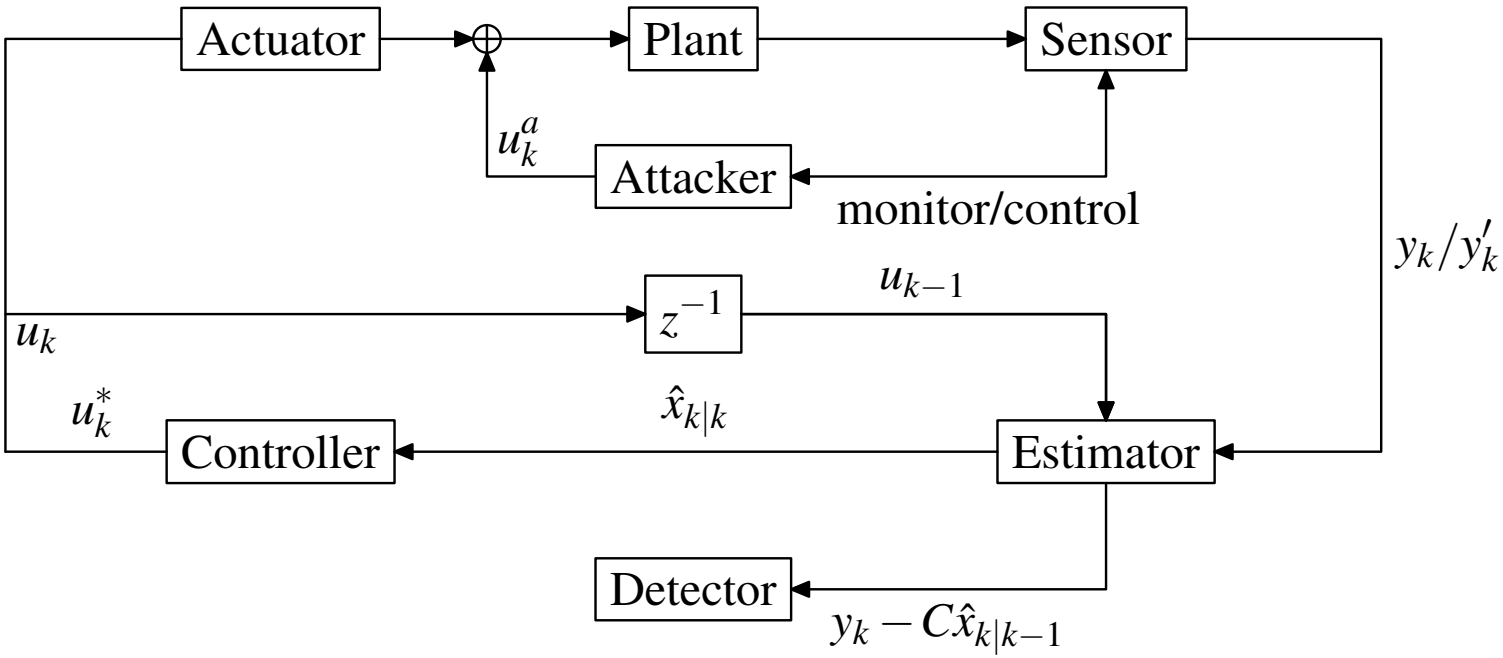
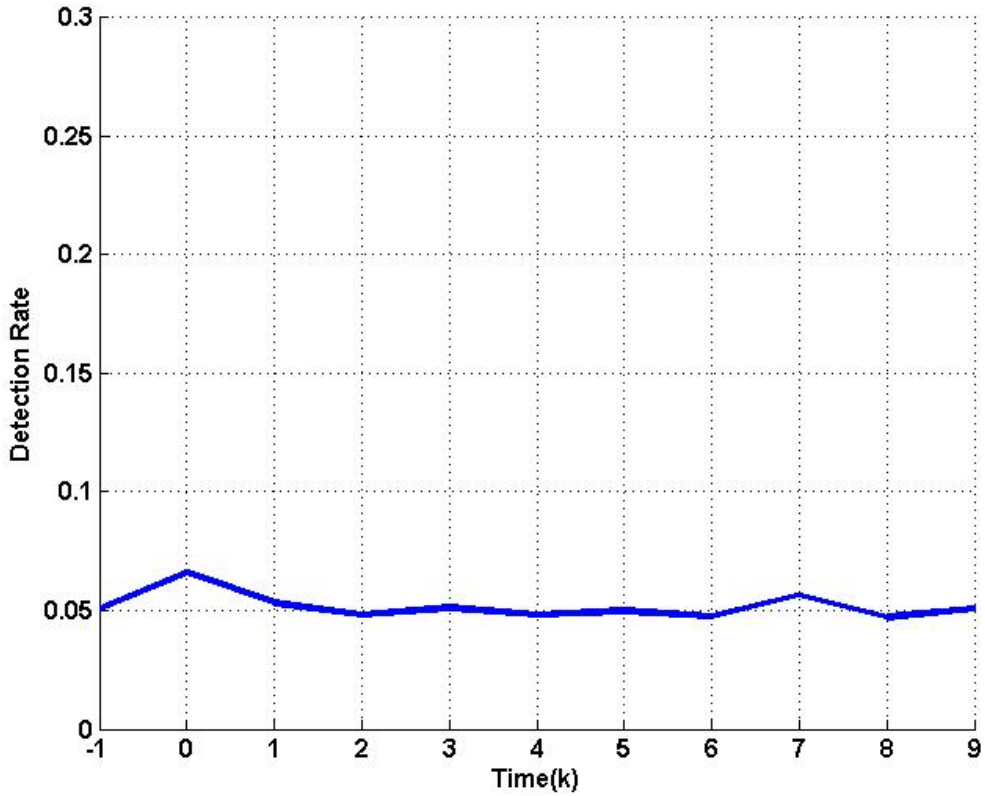


Figure: System Diagram

Simulation

Suppose the attacker records from time $-T$ and replay begins at time 0.



For some systems, the χ^2 detector cannot distinguish system under replay and system without replay.

Detection of Replay Attack

- When replay begins, the update equation of Kalman filter is

$$\begin{aligned}\hat{x}_{k+1|k} &= (A+BL)(I-KC)\hat{x}_{k|k-1} + (A+BL)Ky'_k \\ &= \mathcal{A}\hat{x}_{k|k-1} + (A+BL)Ky'_k\end{aligned}$$

- Since the attacker records from time $-T$, $y'_k = y_{k-T}$. The Kalman filter at time $k-T$ satisfies:

$$\begin{aligned}\hat{x}_{k-T+1|k-T} &= \mathcal{A}\hat{x}_{k-T|k-T-1} + (A+BL)Ky_{k-T} \\ &= \mathcal{A}\hat{x}_{k-T|k-T-1} + (A+BL)Ky'_k.\end{aligned}$$

Detection of Replay Attack

- Hence,

$$\hat{x}_{k|k-1} - \hat{x}_{k-T|k-T-1} = \mathcal{A}^k(\hat{x}_{0|-1} - \hat{x}_{-T|-T-1}).$$

and the innovation $y'_k - C\hat{x}_{k|k-1}$ satisfies

$$\underbrace{y'_k - C\hat{x}_{k|k-1}}_{\text{innovation under replay}} = \underbrace{(y_{k-T} - C\hat{x}_{k-T|k-T-1})}_{\text{innovation without replay}} + \underbrace{C\mathcal{A}^k(\hat{x}_{0|-1} - \hat{x}_{-T|-T-1})}_{\text{converges to 0 if } \|\mathcal{A}\| < 1},$$

- If $\mathcal{A}^k$ converges fast, it is hard to distinguish innovation under replay and innovation without replay!! (They converge in distribution.)

Countermeasures

Replay is feasible because the optimal estimator and controller are deterministic:

$$\hat{x}_{k+1|k} = \mathcal{A}\hat{x}_{k|k-1} + (A + BL)Ky'_k$$

If we add a random control input to the system:

1. If the system respond to this input, then there is no replay attack
2. If the system does not respond, then there is a replay attack
3. Random control inputs act like time stamps
4. Cost: The controller is not optimal any more

Countermeasures

Let control input to be

$$u_k = u_k^* + \Delta u_k,$$

where u_k^* is the LQG control input, Δu_k is a Gaussian random control input with 0 mean and covariance of $\mathcal{Q}$.

Countermeasures

Let control input to be

$$u_k = u_k^* + \Delta u_k,$$

where u_k^* is the LQG control input, Δu_k is a Gaussian random control input with 0 mean and covariance of $\mathcal{Q}$.
Increase in LQG cost: $trace[(U + B^T S B)\mathcal{Q}]$.

Countermeasures

Let control input to be

$$u_k = u_k^* + \Delta u_k,$$

where u_k^* is the LQG control input, Δu_k is a Gaussian random control input with 0 mean and covariance of $\mathcal{Q}$.
Increase in LQG cost: $trace[(U + B^T S B)\mathcal{Q}]$.
Kalman filtering equation:

$$\hat{x}_{k+1|k} = \mathcal{A}\hat{x}_{k|k-1} + (A + BL)Ky'_k + B\Delta u_k$$
$$\hat{x}_{k-T+1|k-T} = \mathcal{A}\hat{x}_{k-T|k-T-1} + (A + BL)Ky'_k + B\Delta u'_k,$$

and

$$y'_k - C\hat{x}_{k|k-1} = y_{k-T} - C\hat{x}_{k-T|k-T-1} + C\mathcal{A}^k(\hat{x}_{0|-1} - \hat{x}_{-T|-T-1})$$
$$+ \boxed{C\sum_{i=0}^{k-1}\mathcal{A}^{k-i-1}B(\Delta u_i - \Delta u_{-T+i})} \leftarrow \text{Can be detected!}.$$

New System Diagram

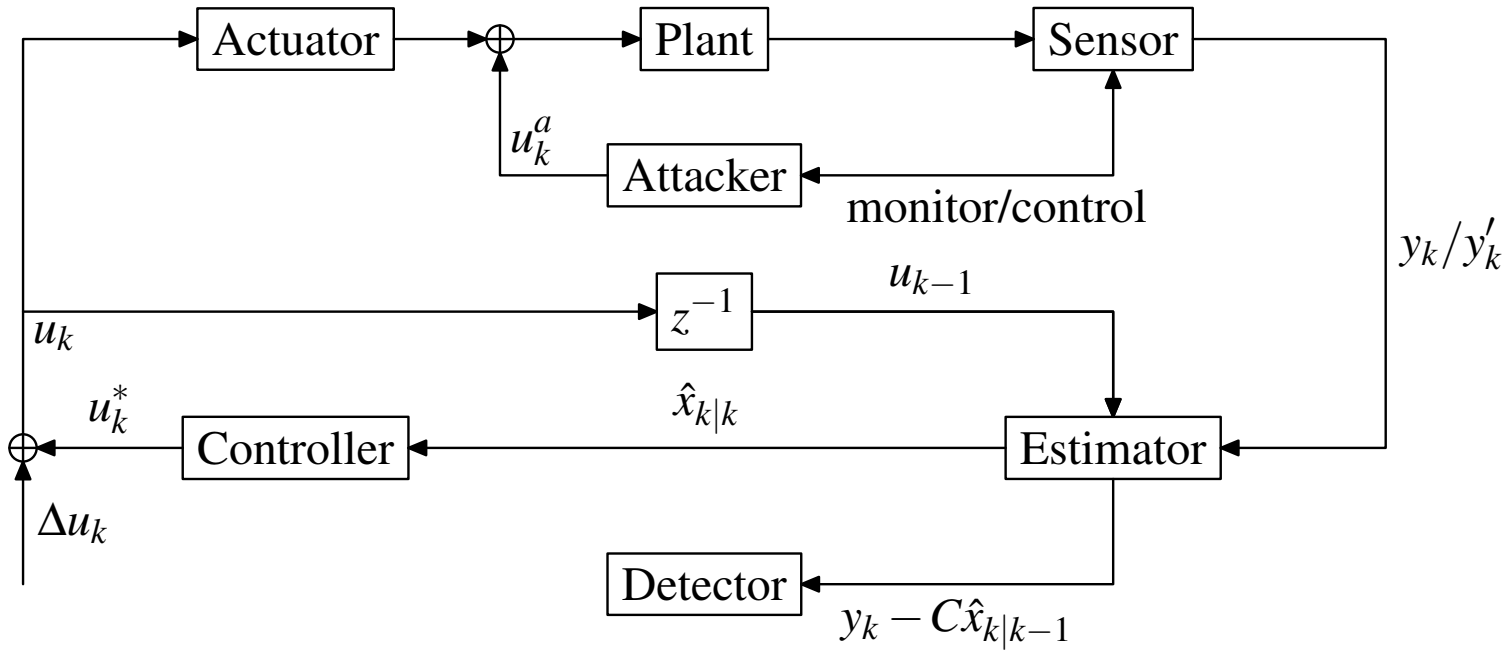


Figure: System Diagram

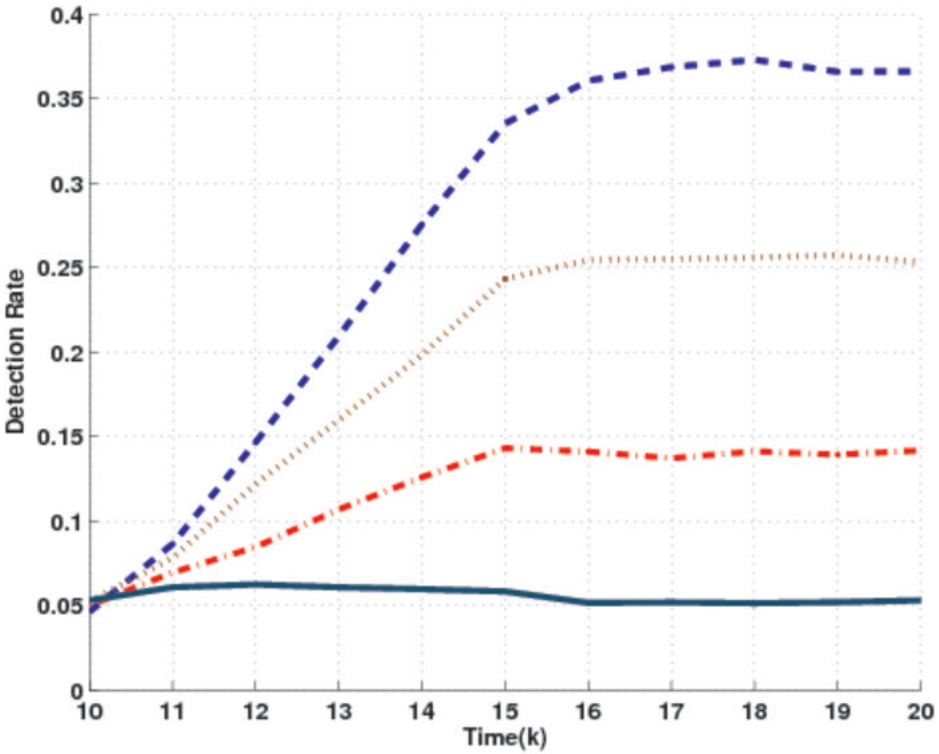
Simulation Result

1. One dimensional system, single sensor

$$x_{k+1} = x_k + u_k + w_k$$
$$y_k = x_k + v_k$$

2. Parameters:
   - $R = 0.1$, $Q = 1$.
   - $W = U = 1$.
   - Detector window size $\mathcal{T} = 5$, false alarm rate 5%.
   - $K = 0.9161$, $L = -0.6180$, $\mathcal{A} = 0.0321$.

Simulation Result



- Blue: $\mathcal{Q} = 0.6$
- Brown: $\mathcal{Q} = 0.4$
- Red: $\mathcal{Q} = 0.2$
- Dark Blue: $\mathcal{Q} = 0$

Figure: Detection Rate of Different Random Signal Strength

Conclusion

While infrastructure is getting smarter, it also becomes more vulnerable. We need to develop a methodology to

- Model attacks
- Analyze the effects of an attack on CPS
- Develop countermeasures to ensure: graceful degradation, reconfiguration, remediation.