# False Data Injection Attacks in Control Systems

Yilin Mo, Bruno Sinopoli

Department of Electrical and Computer Engineering,
Carnegie Mellon University

First Workshop on Secure Control Systems

# Control Systems

- Control Systems are ubiquitous.
- Typical applications of control systems include aerospace, chemical processes, civil infrastructure, energy and manufacturing.
- Many of them are safety-critical.
- Advances in computation and communication technology have greatly increased the capability of control systems. But new challenges arise as the systems become more and more complicated.
- Our goal: analysis and design of secure control systems.

# System Model

We consider the control system 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 state)

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

- The LQG controller minimizes the following cost

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

- The solution is a fixed gain controller

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

where

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

# $\chi^2$ Failure Detector

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

### $\chi^2$ Detector

The $\chi^2$ detector triggers an alarm based on the following event:

$$g_k = (y_k - C\hat{x}_{k|k-1})^T \mathcal{P}^{-1} (y_k - C\hat{x}_{k|k-1}) > \mathit{threshold}.$$

# Attack Model

We assume the following:

1. The attacker knows matrices $A$, $C$, $K$.
2. The attacker can control the readings of a subset of sensors. Hence, the measurement received by the Kalman filter can be written as

$$y_k' = Cx_k' + v_k + \Gamma y_k^a,$$

   where $y_k^a$ is the bias introduced by the attacker, $\Gamma = diag(\gamma_1, \dots, \gamma_m)$ is the sensor selection matrix. $\gamma_i = 1$ if the attacker can control the readings of sensor $i$. $\gamma_i = 0$ otherwise.
3. The attack begins at time 0.
4. The sequence of attacker's inputs $(y_0^a, \dots, y_k^a)$ is chosen before the attack. Hence, $y_k^a$ is independent of $w_k$, $v_k$.

# System Diagram

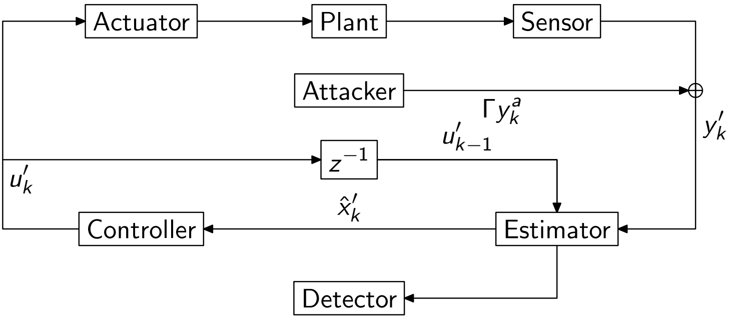


Figure: System Diagram

# Healthy System v.s. Compromised System

### Healthy System

$$
\begin{aligned}
x_{k+1} &= Ax_k + Bu_k + w_k \\
y_k &= Cx_k + v_k \\
z_{k+1} &= y_{k+1} - C(A\hat{x}_k + Bu_k) \\
\hat{x}_{k+1} &= A\hat{x}_k + Bu_k + Kz_{k+1} \\
u_k &= L\hat{x}_k
\end{aligned}
$$

### Compromised System

$$
\begin{aligned}
x'_{k+1} &= Ax'_k + Bu'_k + w_k \\
y'_k &= Cx'_k + v_k + \Gamma y^a_k \\
z'_{k+1} &= y'_{k+1} - C(A\hat{x}'_k + Bu'_k) \\
\hat{x}'_{k+1} &= A\hat{x}'_k + Bu'_k + Kz'_{k+1} \\
u'_k &= L\hat{x}'_k
\end{aligned}
$$

# Difference between the Compromised System and Healthy System

## Dynamics of the Difference

$$\begin{aligned}
\Delta x_{k+1} &= A\Delta x_k + Bu_k, & \Delta z_{k+1} =& \Delta y_{k+1} - C(A\Delta\hat{x}_k + B\Delta u_k), \\
\Delta y_k &= C\Delta x_k + \Gamma y_k^a, & \Delta\hat{x}_{k+1} =& A\Delta\hat{x}_k + B\Delta u_k + K\Delta z_{k+1}, \\
\Delta u_k &= L\Delta\hat{x}_k. & &
\end{aligned}$$

Since $y_k^a$ is independent of $w_k$, $v_k$, we can actually prove that $x_k'$ is Gaussian and

$$E(x_k') = \Delta x_k,\ Cov(x_k') = Cov(x_k).$$

Similar statement is also true for $y_k'$, $z_k'$, $\hat{x}_k'$, $u_k'$. Hence, to characterize the performance of control systems under false data injection attacks, we only need to focus on $\Delta x_k$, $\Delta y_k$, $\Delta z_k$, $\Delta\hat{x}_k$, $\Delta u_k$.

# Successful Attack

### Definition

A sequence of attacker's input $(y_0^a, \ldots, y_N^a)$ is called $\alpha$-feasible if during the attack,

$$D(z_k' \| z_k) = \Delta z_k^T \mathcal{P}^{-1} \Delta z_k / 2 \leq \alpha, \text{ for } \; k = 0, \ldots, N,$$

where $D(z_k' \| z_k)$ is the KL distance between $z_k'$ and $z_k$.

1. It can be proved that the probability of triggering an alarm at time $k$ is an increasing function of $D(z_k' \| z_k)$.
2. If $\alpha$ goes to 0, then the compromised system and healthy system are undistinguishable by the $\chi^2$ detector.

# Constrained Control Problem

1. Under the requirement that $\Delta z_k^T \mathcal{P}^{-1} \Delta z_k / 2 \leq \alpha$, the action of the attacker can be formulated as a constrained control problem, where $y_k^a$ is the input from the attacker.
2. To characterize the resilience of control system, we need to compute the reachable region $R_k$ of $\Delta x_k$.
3. In this talk, we will focus on finding a necessary and sufficient condition under which the union of all $R_k$ is unbounded, i.e. there exists an $\alpha$-feasible attack sequence that can push $\Delta x_k$ arbitrarily far away from 0.

# Main Result

**Theorem**

*$\bigcup_{k=1}^{\infty} R_k$ is unbounded if and only if $A$ has an unstable eigenvalue and the corresponding eigenvector $v$ satisfies:*

1. *$Cv \in span(\Gamma)$, where $span(\Gamma)$ is the column space of $\Gamma$.*
2. *$v$ is in the reachable space of the pair $(A - KCA, K)$.*

1. To check the resilience of control system, one can find all the unstable eigenvector of $A$ and compute $Cv$.
2. If $Cv$ is sparse, then the attacker only need to compromise a few sensors to launch an attack along the direction $v$.
3. To improve the resilience, the defender could add redundant sensors to measure every unstable mode.

# Illustrative Example

We consider a vehicle moving along the $x$-axis, which is monitored by a position sensor and velocity sensor.

## System Description

$$\begin{bmatrix} \dot{x}_{k+1} \\ x_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \dot{x}_k \\ x_k \end{bmatrix} + \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} u_k + w_k,$$

$$y_{k,1} = \dot{x}_k + v_{k,1},$$

$$y_{k,1} = x_k + v_{k,2}.$$

We assume that $Q = R = W = I_2,\ U = 1.$ The Kalman gain and LQG control gain are

$$K = \begin{bmatrix} 0.5939 & 0.0793 \\ 0.0793 & 0.6944 \end{bmatrix},\ L = \begin{bmatrix} -1.0285 & -0.4345 \end{bmatrix}.$$

# Illustrative Example

- It is easy to check the only unstable eigenvector is $v = [0,\ 1]^T$.
- If the position sensor is compromised, then the attacker could push the state $x_k$ to infinity.
- If only the velocity sensor is compromised, then $\bigcup_{k=1}^{\infty} R_k$ is bounded. Here we use an ellipsoidal approximation to compute the inner and outer approximation of $\bigcup_{k=1}^{\infty} R_k$.

# Position Sensor is Compromised

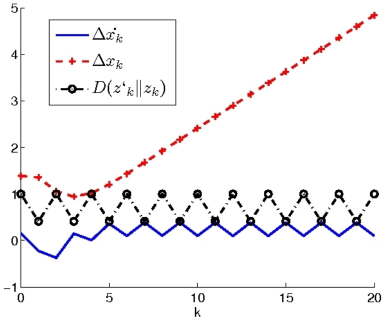


Figure: Evolution of $\Delta\dot{x}_k$, $\Delta x_k$ and $D(z'_k\|z_k)$

# Velocity Sensor is Compromised

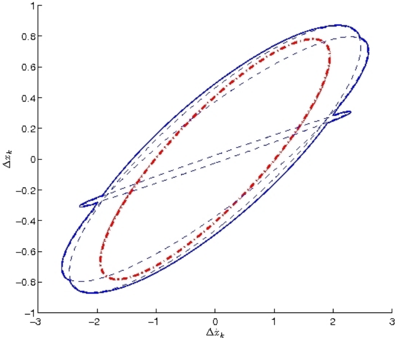


Figure: Inner and Outer Approximation of Reachable Region $\bigcup_{k=1}^{\infty} R_k$ under Constraint $D(z_k' \| z_k) \leq 1$

# Conclusion

In this presentation, we consider the false data injection attacks in control systems.

- We define the false data injection attack model.
- We formulate the action of the attacker as a constrained control problem.
- We prove an algebraic condition under which the attacker could successfully destabilize the system.
- We give a design criterion to improve the resilience of control systems against such kind of attacks.