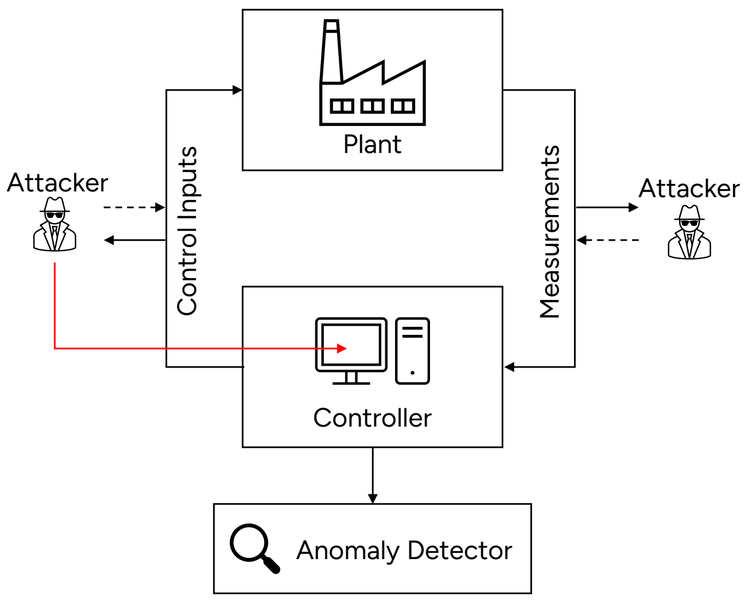
Confidentiality in Networked Control Systems

Enno Breukelman and Henrik Sandberg
{cebre, hsan}@kth.se

WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM



Motivation: Security of Cyber-Physical Systems (CPS)

Ubiquitous use of CPS makes them vulnerable to cyber-attacks, due to communication of control inputs and measurements over a network.

Traditionally, Cyber-attacks are classified by:
Confidentiality, **I**ntegrity, **A**vailability.[1]

Tailored to CPS is the so-called *attacks space*, which is spanned by:
Model Knowledge, Disclosure, and Disruption Resources.[2]

We focus on **stealthy actuator attacks**: the attacker wants to stay undetected.
→ Attacker needs to fool the anomaly detector.
→ This requires knowledge of private information within the controller.

Problem: Confidentiality attack on the controller

The attacker uses $u[k]$ to estimate the controller states $x_c[k]$.

Estimate x_c by $\hat{x}_c$ **perfectly**:

1. Without bias:
   $\mathbb{E}[x_c[k] - \hat{x}_c[k]] = \mathbb{E}[e[k]] = 0$
2. Zero steady-state error covariance:
   $\lim_{k\to\infty} \mathbb{E}[e[k]e[k]^\top] = 0$

Plant

$$x[k+1] = Ax[k] + Bu[k] + w[k]$$
$$y[k] = Cx[k] + v[k]$$

$\uparrow u[k]$ $\qquad$ $y[k] \downarrow$

Controller

$$x_c[k+1] = A_c x_c[k] + B_c y[k]$$
$$u[k] = C_c x_c[k] + D_c y[k]$$

Including independent, zero-mean, Gaussian process and measurement noise $w[k] \sim \mathcal{N}(0, \Sigma_w)$ and $v[k] \sim \mathcal{N}(0, \Sigma_v)$.

Attack strategy #1: Kalman filter

First, let $z[k] = \begin{bmatrix} x[k] \\ x_c[k] \end{bmatrix}$ for the **closed loop state space** formulation.

Due to the noise, this is a random variable. We can set up the Kalman filter as an *unbiased estimator*:

$$z\big[k+1 \mid \{u[i]\}_{i=0}^{k}\big] \sim \mathcal{N}(\hat{z}[k+1], \Sigma_z[k+1]).$$

For the attacker to estimate $x_c[k]$ **perfectly**, we need:

$$\Sigma_z[k] = \mathbb{E}\big[z[k]z[k]^\top\big], \qquad \lim_{k\to\infty} \Sigma_z[k] = \Sigma_\infty = \begin{bmatrix} P & 0 \\ 0 & 0 \end{bmatrix}, \quad P \succeq 0.$$

We can show
1. that Σ_∞ is the *unique* and *strong* solution of the filter's Riccati eq.
2. exponential convergence of the covariance towards Σ_∞.

Attack strategy #2: Unknown Input Observer (UIO)

Describe a series of control signals from a series of measurements:

$$u[k:k+L] = \mathcal{O}_L x_c[k] + \mathcal{J}_L y[k:k+L],$$

with recursively defined observability $\mathcal{O}_L$ and invertibility matrix $\mathcal{J}_L$:

$$\mathcal{O}_L = \begin{pmatrix} C_c \\ \mathcal{O}_{L-1}A_c \end{pmatrix}; \qquad \mathcal{J}_L = \begin{pmatrix} D_c & 0 \\ \mathcal{O}_{L-1}B_c & \mathcal{J}_{L-1} \end{pmatrix}.$$

With parameters E and F, the UIO estimates $x_c[k]$:

$$\hat{x}_c[k+1] = E\hat{x}_c[k] + Fu[k:k+L].$$

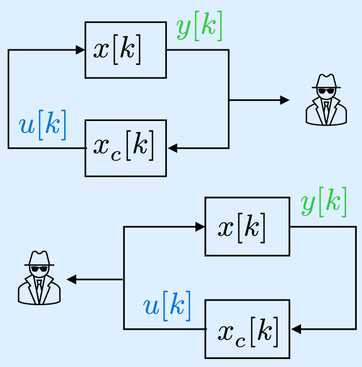
In contrast to standard linear Luenberger observer:
- **independent** of plant outputs $y[k:k+L]$,
- with potential system inherent delay L.

Similarities and Differences

	#1: **Kalman filter**	#2: **UIO**
Estimator	$\hat{z}[k+1] = f_{\mathrm{KF}}(\hat{z}[k], u[k])$	$\hat{x}_c[k+1] = f_{\mathrm{UIO}}(\hat{x}_c[k], u[k:k+L])$
Information	controller, plant, noise	only controller
Closed loop	required to be stable	independent from plant
Matrix conditions	D_c full rank and $\rho(A_c - B_c D_c^\dagger C_c) < 1$	$\operatorname{rank}\begin{bmatrix} A_c - zI & B_c \\ C_c & D_c \end{bmatrix} = n_c + n_y,$ $\forall z \in \mathcal{C}, \lvert z\rvert \geq 1.$
Interpretation	Stable **instantaneous right inverse**.	Stable **L-delay left-inverse** / strongly detectable.

⇒ In both cases: **The controller may not have unstable zeros!**

Conclusion and Outlook



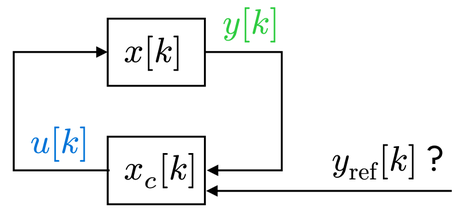
contact:

Sensor attacks (previous work [3]):
Using a **Kalman filter** and plant measurements $y[k]$ for a confidentiality attack, requires the **controller to have stable poles**.

Actuator attacks (our work):
Using a **Kalman filter** and control inputs $u[k]$ for a confidentiality attack, requires the **controller to have stable zeros**, and D_c full rank.

Much less restrictive is the use of an **Unknown Input Observer**:
👍 requires less model knowledge and merely strong detectability.
👎 inherits a potential delay, convergence speed dependent on zeros.

Future work
Estimate an unknown reference to the controller.



[1]: M. Bishop, E. Sullivan, and M. Ruppel, "Computer security: art and science", [2]: André Teixeira, Iman Shames, Henrik Sandberg, Karl Henrik Johansson, "A secure control framework for resource-limited adversaries", [3]: Umsonst, David; Sandberg, Henrik, "On the confidentiality of controller states under sensor attacks"