瀋陽航空航天大學
SHENYANG AEROSPACE UNIVERSITY

Watermark-based Proactive Defense Strategy Design For Cyber-Physical Systems With Unknown-but-bounded Noises

Liu Hao

Shenyang Aerospace University, China

Hao Liu, Yuzhe Li, Qing-Long Han, Tarek Raissi. Watermark-based proactive defense strategy design for cyber-physical systems with unknown-but-bounded noises, IEEE Transactions on Automatic Control, 2022, doi: 10.1109/TAC.2022.3184396.

Outline

1. Introduction
2. Attack detection and performance analysis
3. The design of functions and watermarks
4. Attack detection in different scenarios
5. Conclusions and future work

Introduction

- In recent years, cyber attacks which may deteriorate the system performance have attracted much attention of researchers.
- The goal of Man-In-The-Middle (MITM) attacks is to deteriorate the performance of state estimation by corrupting sensors' data, while attempting to remain stealthy.
- Existing proactive detection approaches based on watermarking can only be utilized to detect malicious attacks on the system with Gaussian noises, and cannot be directly employed to detect attacks for the system with unknown-but-bounded noises.

Introduction _ System model

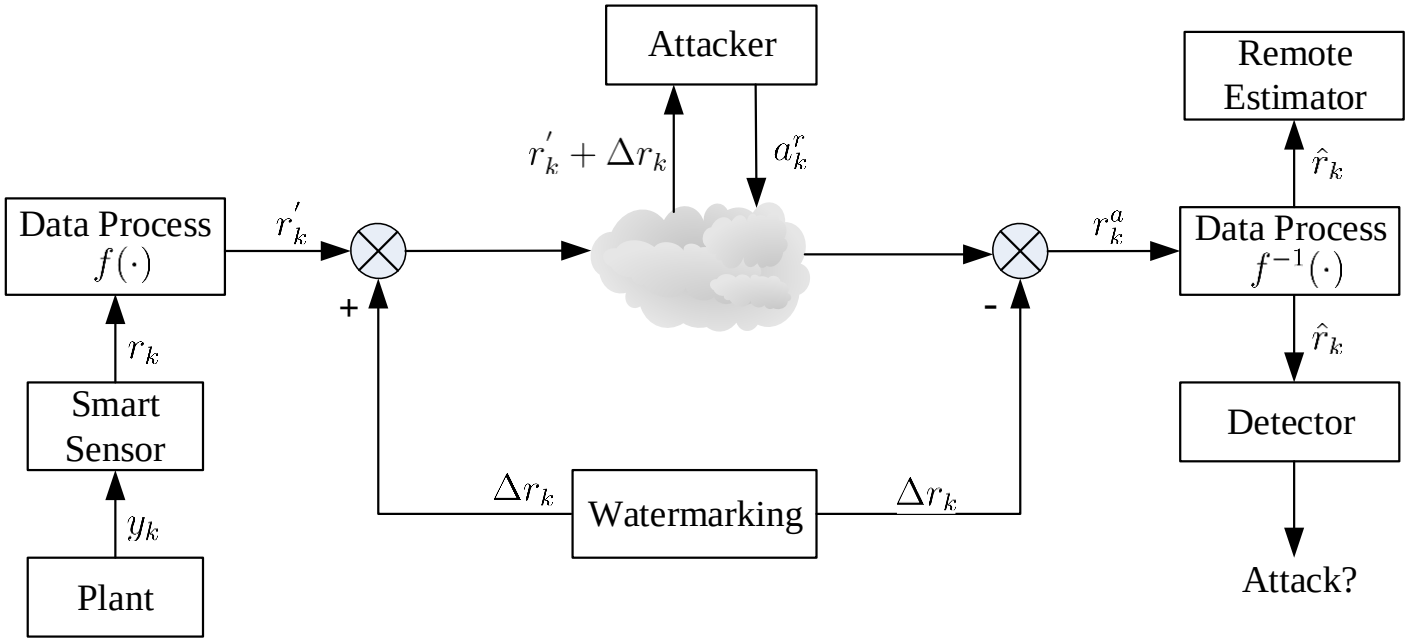


Fig. 1 System architecture under watermark-based proactive defense strategy

Introduction _ System model

System model:

$$x_{k+1} = Ax_k + \omega_k \qquad \omega_k \in \mathcal{W} = \langle 0, H_\omega \rangle$$
$$y_k = Cx_k + v_k \qquad v_k \in \mathcal{V} = \langle 0, H_v \rangle$$

The smart sensor and remote estimator:

$$\hat{x}_{k+1} = A\hat{x}_k + L(y_k - C\hat{x}_k)$$

Data processes:

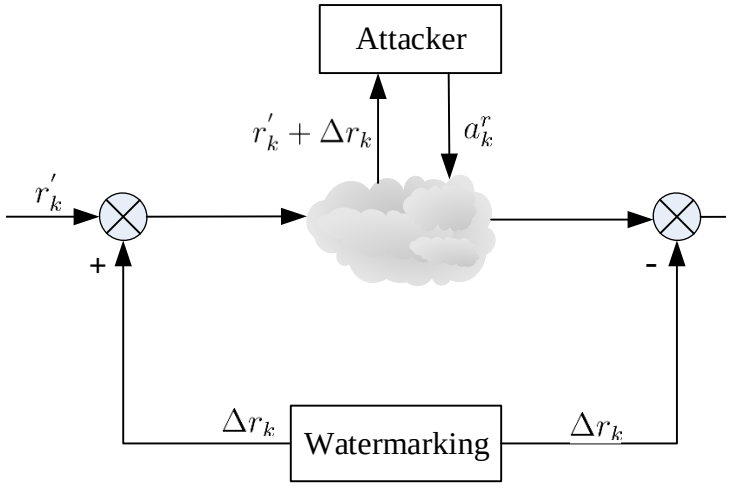
$$r'_k = f(r_k) \qquad r_k = y_k - C\hat{x}_k \qquad f(r_k) \triangleq \begin{bmatrix} f(r_k^1) \\ \vdots \\ f(r_k^{n_y}) \end{bmatrix}$$

Assumption: *f* is a continuous and monotone function and its invertible function exists.

Introduction _ Attack model

Attack model: $a_k^r = g\left(r_k' + \Delta r_k\right)$

$$g(\bar{r}_k) \triangleq \begin{bmatrix} g(\bar{r}_k^1) \\ \vdots \\ g(\bar{r}_k^{n_y}) \end{bmatrix}, \quad \bar{r}_k = r_k' + \Delta r_k.$$



It is assumed that the function $g(\cdot)$ is continuous.

The attack detector:

$$\begin{cases} \hat{r}_k^i \in [\underline{r}_i, \bar{r}_i], & \text{Attack-free} \\ \hat{r}_k^i \notin [\underline{r}_i, \bar{r}_i], & \text{Attacked} \end{cases}$$

$$\bar{r} = \max_{k \geq 0} \left\{ p_k^r + \sum_{j=1}^{s} |(H_k^r)_{ij}| \right\}$$

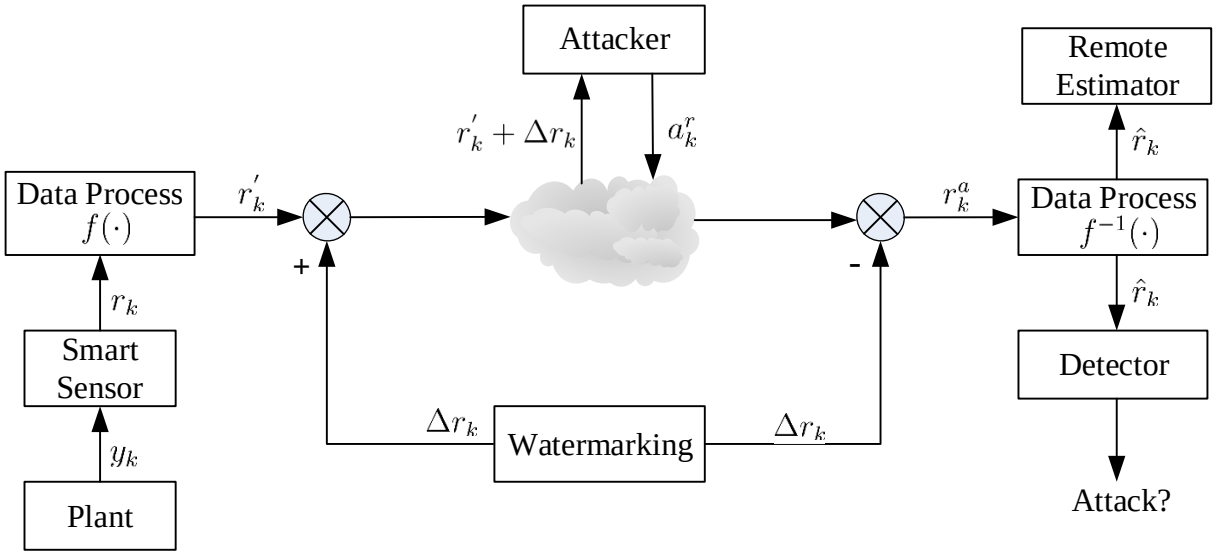
$$\underline{r} = \min_{k \geq 0} \left\{ p_k^r - \sum_{j=1}^{s} |(H_k^r)_{ij}| \right\}$$

Introduction _ Problems to be solved

Problem 1: How to design the function f and the watermark such that the proactive detection rate can reach 100%?

Problem 2: How to analyze the effect on the estimation performance of the false-data-injection (FDI) attacks under the proposed proactive defense strategy?

Attack detection and performance analysis



$$\hat{r}_k = f^{-1}\left(f(r_k) + g(f(r_k) + \Delta r_k)\right) \in [\zeta_k^1,\ \zeta_k^2]$$

$$\zeta_k^1 = \min\left\{f^{-1}(\delta_k^1 + \xi_k^1), f^{-1}(\delta_k^2 + \xi_k^2)\right\},\ \zeta_k^2 = \max\left\{f^{-1}(\delta_k^1 + \xi_k^1), f^{-1}(\delta_k^2 + \xi_k^2)\right\}$$

$$\xi_k^1 = \min\left\{g(\delta_k^1 + \Delta r_k^1), g(\delta_k^2 + \Delta r_k^2)\right\},\ \xi_k^2 = \max\left\{g(\delta_k^1 + \Delta r_k^1), g(\delta_k^2 + \Delta r_k^2)\right\}$$

$$\delta_k^1 = \min\left\{f\left(r_k^1\right), f\left(r_k^2\right)\right\},\ \delta_k^2 = \max\left\{f\left(r_k^1\right), f\left(r_k^2\right)\right\}$$

$$\Delta r_k \in [\Delta r_k^1, \Delta r_k^2] \quad \Delta r_k^1 = p_k^{\Delta r} - \sum\nolimits_j \left|(H_k^{\Delta r})_{ij}\right|,\ \Delta r_k^2 = p_k^{\Delta r} + \sum\nolimits_j \left|(H_k^{\Delta r})_{ij}\right|$$

Attack detection

Theorem 1: Consider the cyber-physical system with UBB noises, which is shown in Fig. 1. If the system is equipped with the proposed attack detector and the designed watermark, then the detection rate for i^{th} channel can be calculated as follows:

Case 1: Attacks are stealthy if $[\zeta_{k,i}^1,\ \zeta_{k,i}^2] \subseteq [\underline{r}_i,\ \bar{r}_i]$ i.e., the detection rate is zero at time step *k*.

Case 2: If $\zeta_{k,i}^1 \geq \bar{r}_i$ or $\zeta_{k,i}^2 \leq \underline{r}_i$, then the detection rate is 100% at time step *k*, i.e.,

$$\Pr\{\hat{r}_k^i \notin [\underline{r}_i,\ \bar{r}_i]\} = 1$$

Case 3: If $\underline{r}_i \leq \zeta_{k,i}^1 < \bar{r}_i < \zeta_{k,i}^2$, then the detection rate at time step *k* is

$$\Pr\{\hat{r}_k^i \notin [\underline{r}_i,\ \bar{r}_i]\} = \frac{\zeta_{k,i}^2 - \bar{r}_i}{\zeta_{k,i}^2 - \zeta_{k,i}^1}$$

Case 4: If $\zeta_{k,i}^1 \leq \underline{r}_i < \zeta_{k,i}^2 < \bar{r}_i$, then the detection rate at time step *k* is

$$\Pr\{\hat{r}_k^i \notin [\underline{r}_i,\ \bar{r}_i]\} = \frac{\bar{r}_i - \zeta_{k,i}^1}{\zeta_{k,i}^2 - \zeta_{k,i}^1}$$

Performance analysis

Definition: (Optimal Watermark) The proposed watermark Δr_k is optimal if it is designed such that the condition $\Pr\{\hat{r}_k^i \notin [\underline{r}_i,\ \bar{r}_i]\} = 1$ holds for any $k \in \mathbb{N}_{\geq 1}$ and $i = 1, 2, \ldots, n_y$.

Theorem 2: Consider the cyber-physical system with UBB noises, which is shown in Fig. 1. If the system is equipped with the proposed attack detector and the designed watermark, then the state estimation error e_k satisfies the condition

$$e_k \in [\varrho_k^1, \varrho_k^2]$$

$$\varrho_k^1 = \sum_{l=0}^{k-1} \left(-A^{k-l-1}L\right) \zeta_l^1 - \sum_j \left|(A^k H_0)_{ij}\right| - \sum_{l=0}^{k-1} \sum_j \left|(A^l H_\omega)_{ij}\right|,$$

$$\varrho_k^2 = \sum_{l=0}^{k-1} \left(-A^{k-l-1}L\right) \zeta_l^2 + \sum_j \left|(A^k H_0)_{ij}\right| + \sum_{l=0}^{k-1} \sum_j \left|(A^l H_\omega)_{ij}\right|.$$

The design of function *f*

The design of $f(\cdot)$

According to Assumption 2, *f* is a continuous and monotone function and its invertible function exists. Therefore, *f* can be designed in a variety of forms. For example, we can choose

$$f(r_k) = r_k^3 \qquad f(r_k) = \exp\{r_k\}$$

For simplicity, a linear function is designed, which satisfies

$$f(r_k) = \alpha_k r_k$$

where $\alpha_k \in \mathbb{R}$ is a time-varying parameter and $\alpha_k \neq 0$.

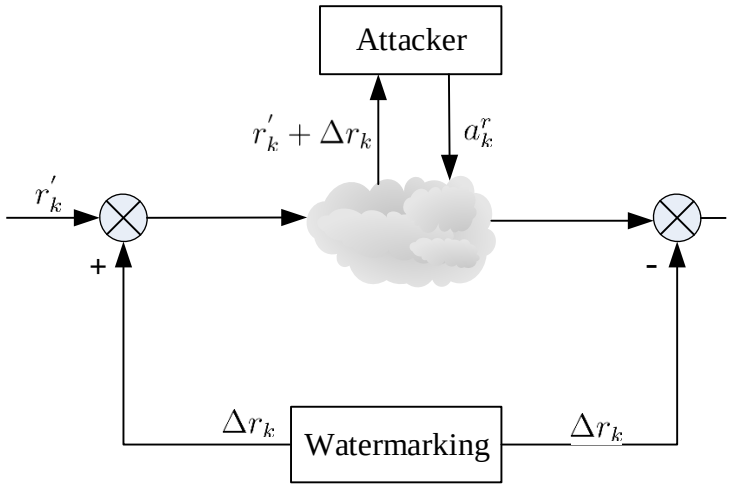
$$f^{-1}(r_k^a) = \frac{1}{\alpha_k} r_k^a$$

The design of attack function

The design of $g(\cdot)$

$$a_k^r = g(f(r_k) + \Delta r_k) = \Gamma_k(\alpha_k r_k + \Delta r_k) + \mu_k$$

where $\Gamma_k \in \mathbb{R}^{n_y \times n_y}$ and $\mu_k \in \mathbb{R}^{n_y}$ are parameters determined by an attacker.



The design of watermark

The design of Δr_k

Theorem 3: Consider the cyber-physical system with UBB noises, which is shown in Fig. 1 and the watermark $\Delta r_k \in \langle p_k^{\Delta r}, H_k^{\Delta r} \rangle$. The detection rate can reach 100% if $\Gamma_k = -I_{n_y}$, $\mu_k = 0$ and the watermark satisfies the following condition:

$$-p_k^{\Delta r} - \sum_j \left| \left(H_k^{\Delta r} \right)_{ij} \right| > \alpha_k \bar{r}$$

or

$$-p_k^{\Delta r} + \sum_j \left| \left(H_k^{\Delta r} \right)_{ij} \right| < \alpha_k \underline{r}$$

where $\alpha_k \in \mathbb{R}_{>0}$.

Attack detection _ Case 1

Case 1: Both α_k and Δr_k are not available to an attacker

MITM attack can be designed as

$$a_k^r = -(\alpha_k r_k + \Delta r_k) + \Delta r_k^{guess} + \alpha_k^{guess}\mu_k$$

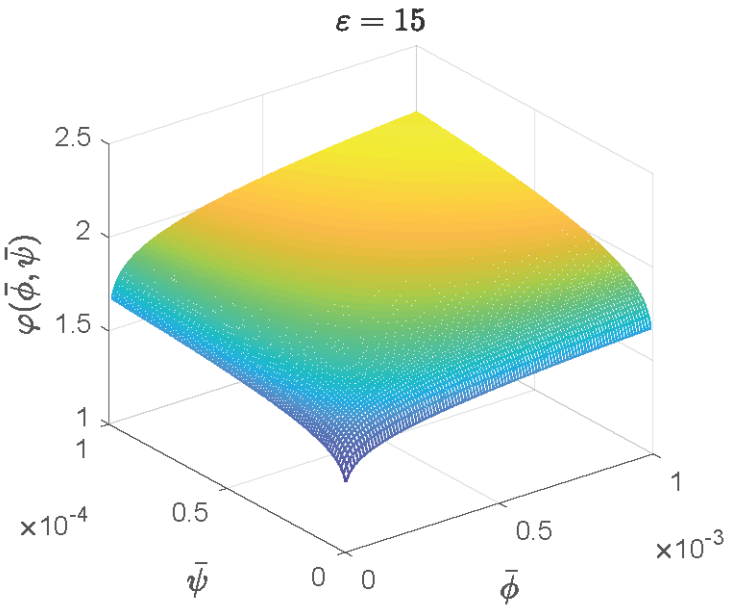
$$\Pr\{\alpha_k^{guess} = \alpha_k\} = \phi_k \leq \bar{\phi}$$

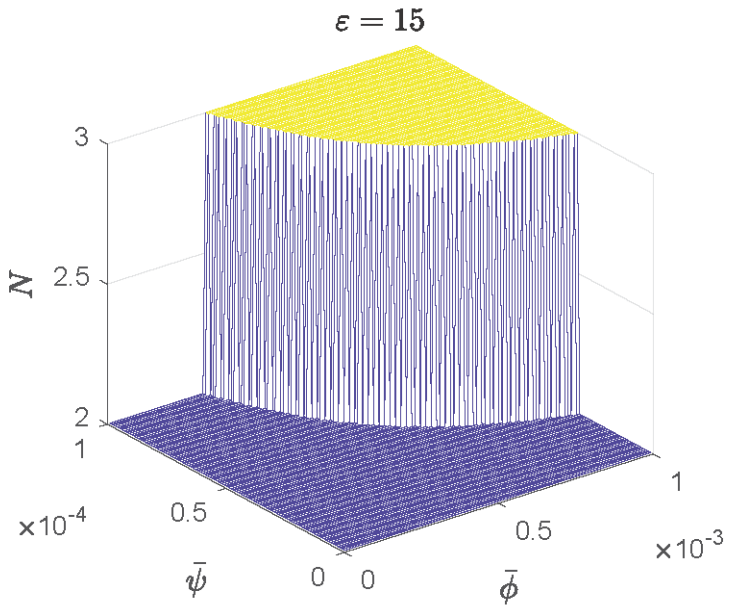
$$\Pr\{\Delta r_k^{guess} = \Delta r_k\} = \psi_k \leq \bar{\psi}$$

$$\prod_{k=1}^{N} \Pr\{\alpha_k^{guess} = \alpha_k, \Delta r_k^{guess} = \Delta r_k\} \to 0$$

$$N = \left\lceil \varepsilon \frac{\ln 0.1}{\ln \bar{\phi} + \ln \bar{\psi}} \right\rceil \qquad \varepsilon \in \mathbb{N}_{>0}$$

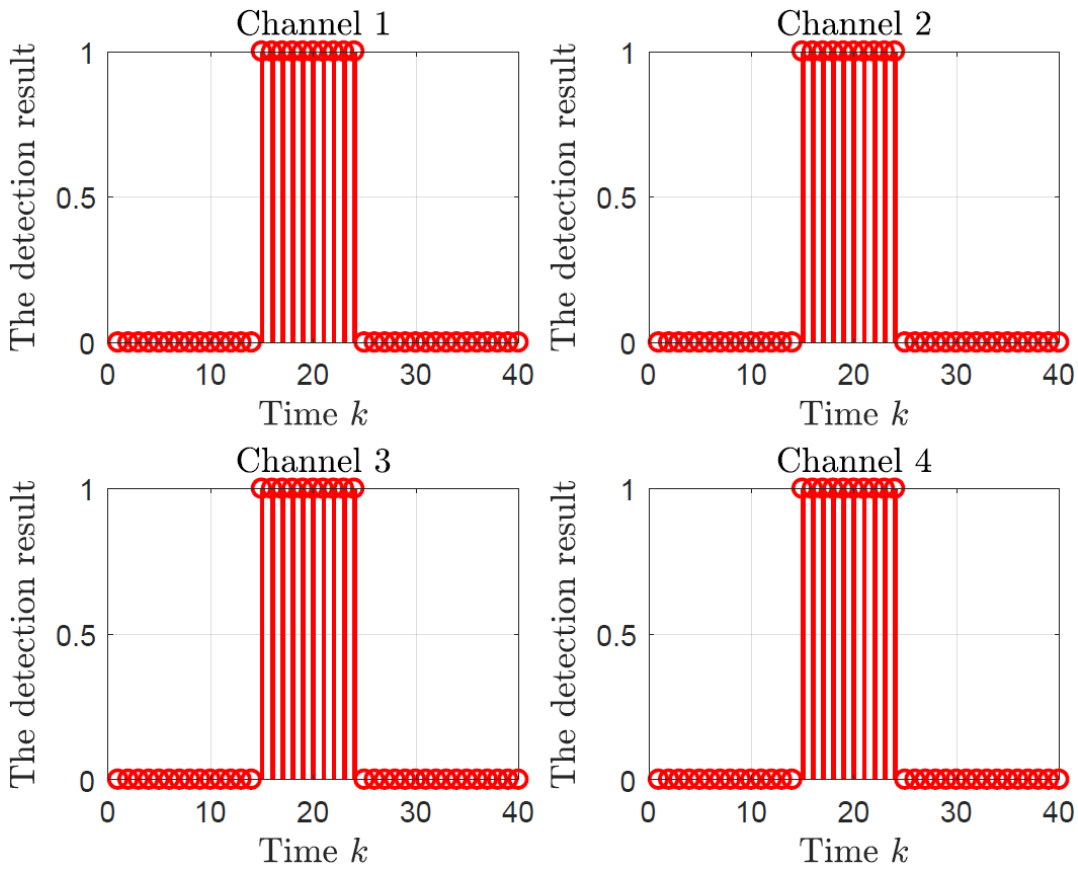
Attack detection _ Case 1





$$\bar{\phi} = 10^{-4},\ \bar{\psi} = 10^{-4},\ \varepsilon = 15 \Rightarrow N = 2$$

Attack detection _ Case 1



The detection results in Case 1

Attack detection _ Case 2

Case 2: Only α_k is available to an attacker

MITM attack can be designed as

$$a_k^r = g(\alpha_k r_k + \Delta r_k) = -\alpha_k r_k + \alpha_k \mu_k$$

$$r_k = y_k - C\hat{x}_k$$

Not available

$$a_k^r = \alpha_k \mu_k \Rightarrow \hat{r}_k = r_k + \mu_k$$

Attack detection _ Case 2

If an attacker guesses the watermark, then

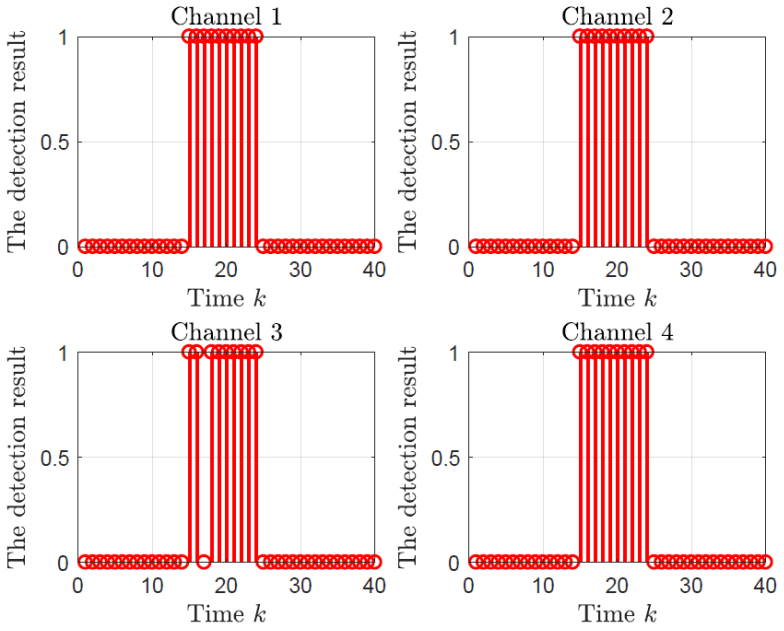
$$a_k^r = -\alpha_k r_k + \Delta r_k - \Delta r_k^{guess}$$

which implies

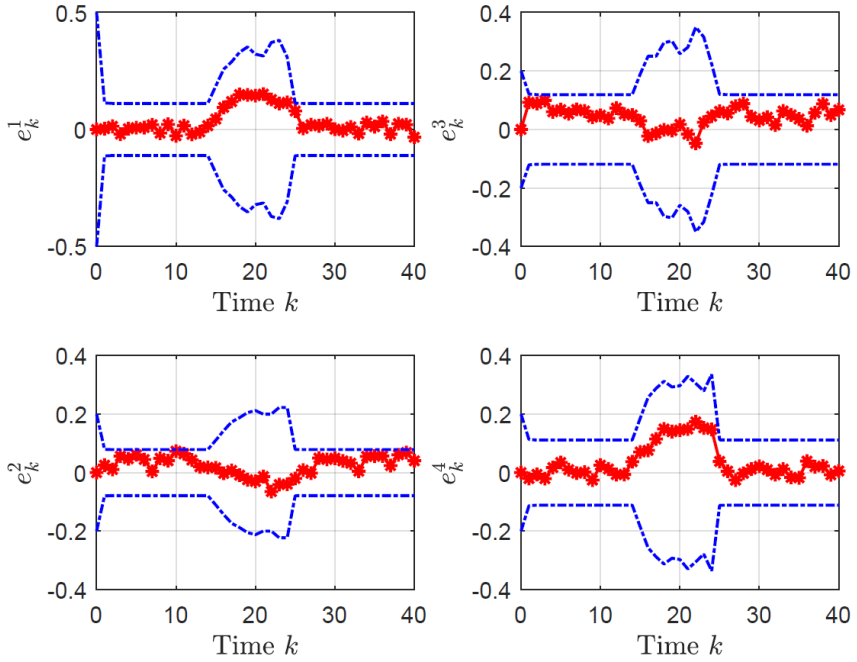
$$\hat{r}_k = \alpha_k^{-1} \left(\Delta r_k^{guess} - \Delta r_k \right)$$

A small α_k corresponds to a large $\hat{r}_k$.

Attack detection _ Case 2

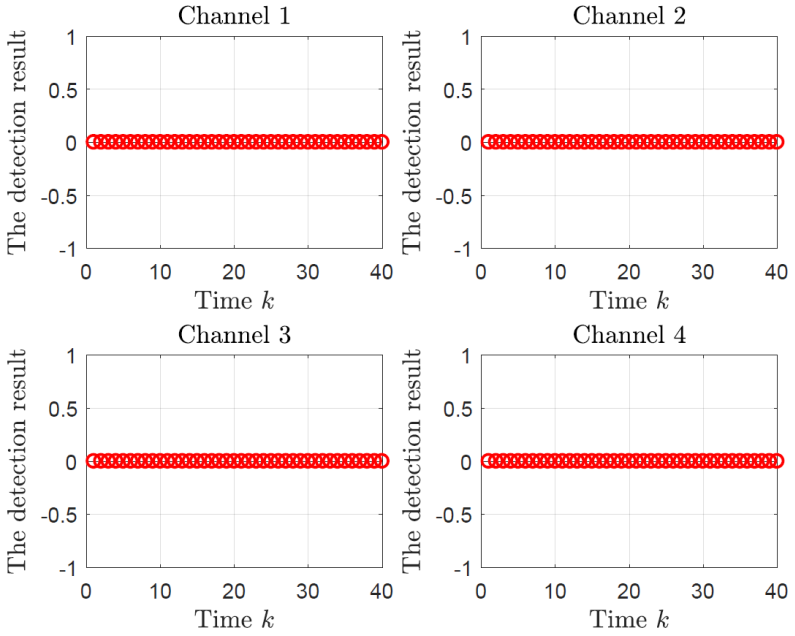


The detection results when $[\underline{\mu}, \bar{\mu}] = [\underline{r}, \bar{r}]$

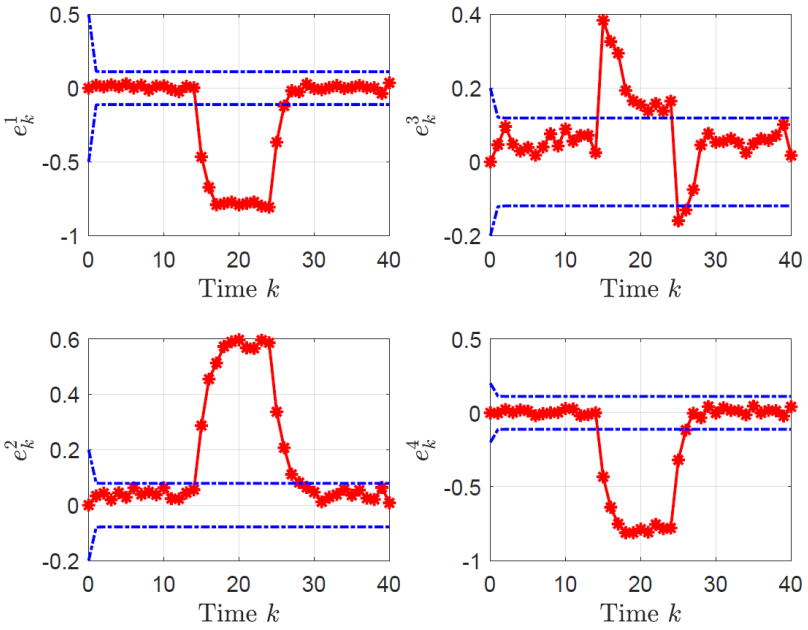


The estimation error $[\underline{\mu}, \bar{\mu}] = [\underline{r}, \bar{r}]$

Attack detection _ Case 2



The detection results when $\underline{\mu} = 0.5\underline{r},\ \bar{\mu} = 0.5\bar{r}$



The estimation error when $\underline{\mu} = 0.5\underline{r},\ \bar{\mu} = 0.5\bar{r}$

Attack detection _ Case 3

Case 3: Only Δr_k is available to an attacker

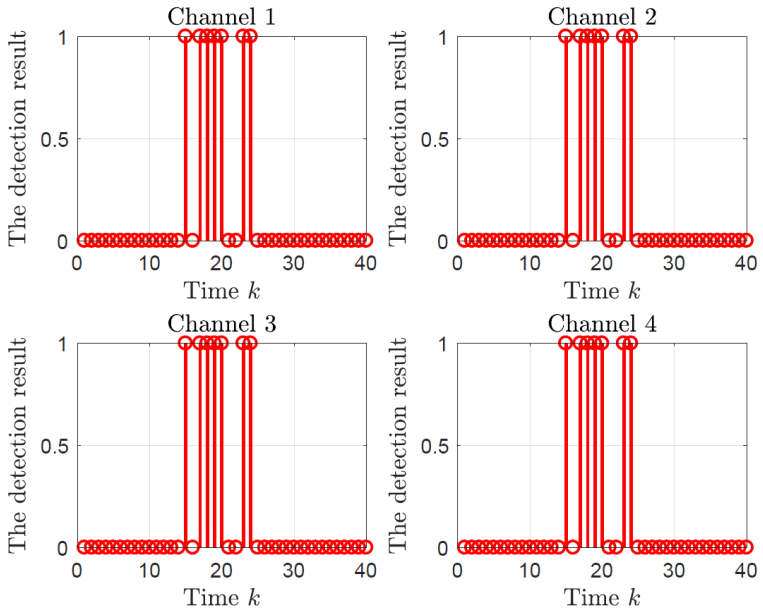
MITM attack can be designed as

$$a_k^r = -\alpha_k r_k + \alpha_k^{guess} \mu_k$$

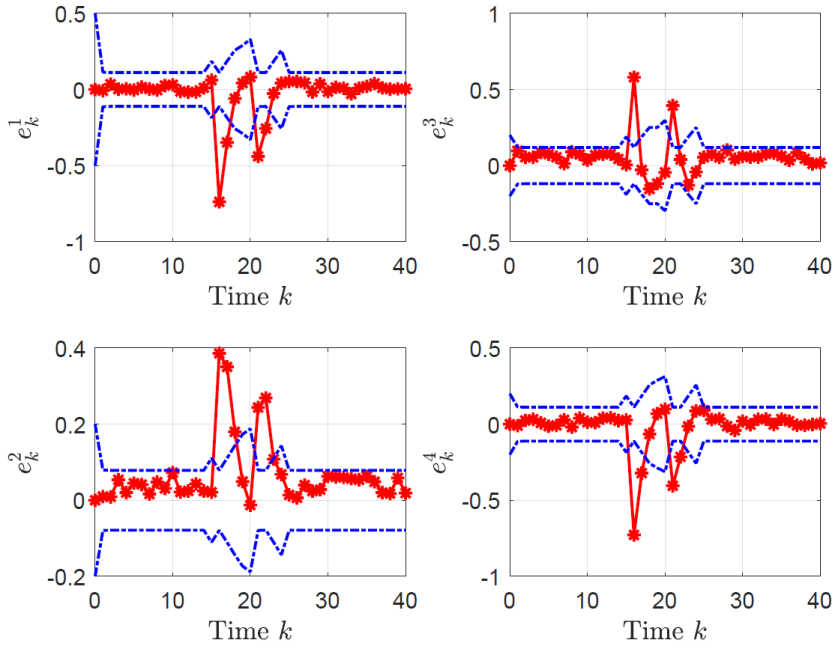
$$\hat{r}_k = \alpha_k^{-1} \alpha_k^{guess} \mu_k$$

$\alpha_k r_k + \Delta r_k$ is available, thus $\alpha_k r_k$ is known.

Attack detection _ Case 3



The detection results when $\mu_k \neq 0$



The estimation error when $\mu_k \neq 0$

Attack detection _ Case 4

Case 4: Both α_k and Δr_k are available to an attacker

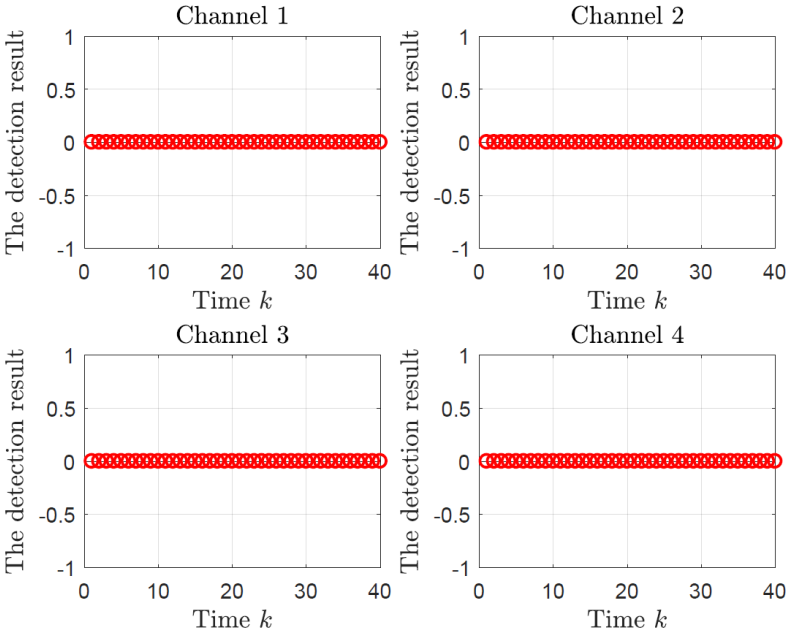
MITM attack can be designed as

$$a_k^r = -\alpha_k(r_k - \mu_k)$$

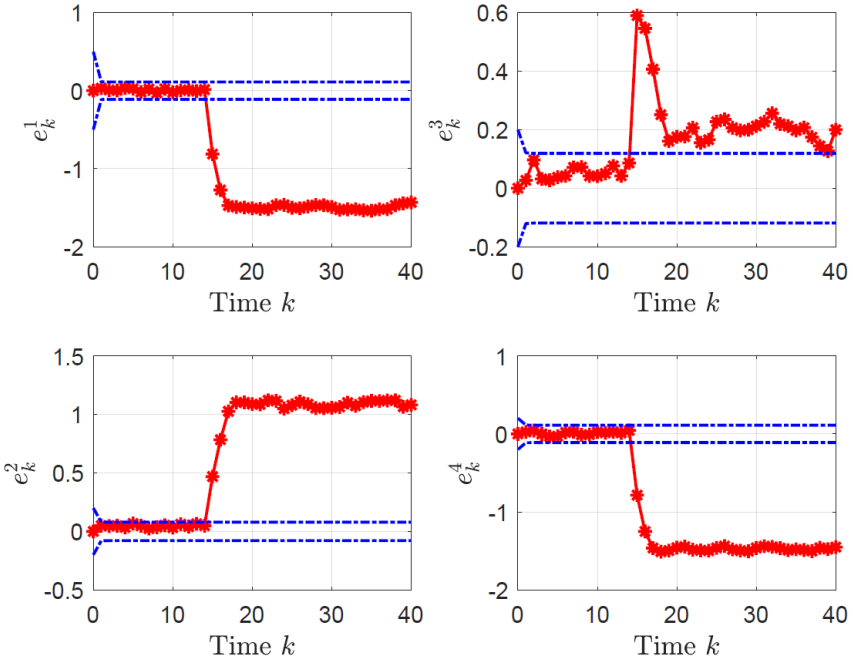
$$\hat{r}_k = \mu_k$$

$\mu_k \in [\underline{r}, \bar{r}]$ → all attacks can remain stealthy.

Attack detection _ Case 4



The detection results



The estimation error

Conclusions and future work

- It is almost impossible for an attacker to launch stealthy MITM attacks if both the parameter α_k and the watermark Δr_k are not available to him or her. This conclusion is also applicable to the case that only the parameter α_k is available to the attacker.
- When the watermark Δr_k can be obtained by an attacker, then he or she can design the corresponding stealthy attacks. Moreover, the estimation error e_k can be further affected when both the parameter α_k and the watermark Δr_k are available to the adversary.
- It is necessary for the defender to choose a small α_k and ensure that both the parameter α_k and the watermark Δr_k are time-varying and secret to an attacker, which can prevent an attacker from continuously launching stealthy attacks.

- In our future work, we will investigate the attack detection issue for the nonlinear CPS with unknown-but-bounded noises. In addition, the UBB noises may be characterized by non-convex sets.

Thank you for listening!

瀋陽航空航天大學
SHENYANG AEROSPACE UNIVERSITY

Watermark-based Proactive Defense Strategy Design For Cyber-Physical Systems With Unknown-but-bounded Noises

Liu Hao

Shenyang Aerospace University, China

Hao Liu, Yuzhe Li, Qing-Long Han, Tarek Raissi. Watermark-based proactive defense strategy design for cyber-physical systems with unknown-but-bounded noises, IEEE Transactions on Automatic Control, 2022, doi: 10.1109/TAC.2022.3184396.