2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Replay Attack Detection for Cyber-Physical Systems with Sensitive States

Tao Chen, Lei Wang, Xiaoqiang Ren, Zhitao Liu and Hongye Su

Abstract— **In cyber-physical systems, replay attacks are a type of deception attacks that involve replaying previously recorded sensor data while injecting attack signals into the physical plant. The replay attacks are difficult to detect because the data being replayed is normal. To detect replay attacks, authentication signals are commonly used. However, injecting authentication signals may affect the performance of system states that are sensitive to noise or disturbance. To address this issue, the effect of authentication signals on sensitive states is measured in the sense of mean square error, and conditions that the authentication signal has a limited effect or even no effect on sensitive states are presented. Combining these conditions, optimization problems are constructed to design the authentication signal. Finally, the proposed method is validated through simulation results.**

I. Introduction

Cyber-physical systems (CPSs) are complex systems that integrate control, communication, and computing technologies, with wide applications in practical industrial systems and products [1], [2]. However, unlike traditional control systems, CPSs rely on remote transmissions subject to attack risks that may cause significant damages [3], [4], [5].

Replay attack is a kind of deception attack that has received significant research attention. It involves hijacking sensors, recording sensor data for a certain amount of time, and replaying the recorded data while injecting arbitrary attack signals into the physical plant. According to [6], the widely used χ^2 detector may fail to detect such attacks since the recorded data is normal, which makes it more harmful.

To detect replay attacks, [6] designed an authentication signal consisting of i.i.d Gaussian noise added to the control signal. They also analyzed the benefits of this method for detecting replay attacks and the performance loss incurred by adding the authentication signal. Based on this research, over the past decade, research on replay attack detection has mainly focused on two aspects: improving detection effectiveness and reducing performance loss caused by authentication signals. For the first aspect, [7] proposed a dynamic watermarking method that depends on the system's dynamics, resulting in an unstable residue signal when a replay attack occurs and achieving a probability of detection equal to one. [8] designed a frequency-based authentication signal to determine which channels are affected by the replay attack. [9] and [10] expanded on the work of [6]. Specifically, [9] formulated optimization problems to maximize detection effectiveness while limiting performance loss, while [10] relaxed the authentication signal to a stationary process generated by a hidden Markov model (HMM). To reduce performance loss caused by authentication signals, [11] proposed a periodic watermarking scheduling approach based on the discontinuous replay attack model, which reduces control costs. [12] used communication error to assist replay attack detection, eliminating the need for an artificial authentication signal. This approach is only suitable for a specific type of CPS involving an additive white Gaussian noise channel. [13] designed a stochastic coding scheme in the sensor-controller channel, where the coding and decoding numbers do not match during a replay attack, allowing for detection. However, this approach requires the dynamic coding and decoding numbers to be known on both the plant and estimator sides, which could be susceptible to attacks.

The authentication signal is essentially a disturbance input. In practice, there may exist states which are sensitive to significant disturbances. For example, in chemical reactions, maintaining a steady temperature is crucial for the success of some reactions [14], [15]. Similarly, in robotics, keeping the lateral direction of a moving robot stable is necessary to maintain tracking accuracy [16]. The front joint of a manipulator should also avoid large disturbances to maintain the accuracy of the end-point [17]. In view of this, the above-mentioned attack detection methods may not be suitable for real systems as they may not adequately protect vulnerable states from being overly disrupted.

In this paper, the problem of designing authentication signals with limited effect on sensitive states in the mean-square sense is studied. By using the mean square error of sensitive states impacted by and free of authentication signals to measure their effect, we establish necessary and sufficient conditions on the authentication signal having a limited effect or even no effect on sensitive states. With these conditions, we further investigate the design of authentication signal by formulating it as optimization problems. The main contributions of this paper lie in establishing necessary and sufficient conditions of authentication signals with limited or even no effect on sensitive states for reply attack detection, and formulating optimization problems to design the optimal authentication signal.

The remainder of the paper is structured as follows. Section II provides preliminary information on replay attacks and our motivation for this study. In Section III, we describe

This research was supported by the National Natural Science Foundation of China under Grant No. 62203386, and Zhejiang Provincial Natural Science Foundation of China under Grant No. LZ23F030008.

Tao Chen, Lei Wang, Zhitao Liu and Hongye Su are with the College of Control Science and Engineering, Zhejiang University, P.R. China. (E-mail: tao_chen; lei.wangzju; ztliu@zju.edu.cn; hysu@iipc.zju.edu.cn).

Xiaoqiang Ren is with the School of Mechatronic Engineering and Automation, Shanghai University, P.R. China.(E-mail:xqren@shu.edu.cn)

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

our method for designing the authentication signal. Simulation results are presented in Section IV, and we conclude this paper in Section V.

Notation. We denote by $\mathbb{R}$ the real numbers, $\mathbb{R}^n$ the real space of n dimension for any positive integer n, $\mathbb{N}$ the set of natural numbers, and $\mathbb{P}^+$ the set of positive semi-definite matrix. For a variable $X \in \mathbb{R}^n$, X_k represents the value of X at time instant k, and $X_{i,k}$ denotes the i^{th} element of X_k. For $A \in \mathbb{R}^{n\times n}$, A^{T} denotes the transpose of matrix A, $\mathrm{Tr}(A)$ denotes the trace of A, and $\rho(A)$ denotes the spectral radius of A. $\mathcal{N}(\mu, \Sigma)$ denotes the Gaussian distribution with mean μ and covariance Σ. I_n denotes identity matrix of dimension $n\times n$. $\mathbb{E}(\cdot)$ denotes the expectation of a random variable. For a set $\mathcal{C}$, $|\mathcal{C}|$ denotes the total number of elements in $\mathcal{C}$.

II. Problem Formulation

A. System Description

Consider the LQG control problem for linear time-invariant physical plant of the form

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

where $x_k \in \mathbb{R}^n$ is the system state, $u_k \in \mathbb{R}^m$ is the control input, $y_k \in \mathbb{R}^p$ is the sensor measurement, $w_k \sim \mathcal{N}(0, Q)$ is the process noise, and $v_k \sim \mathcal{N}(0, R)$ is the measurement noise. Both w_k and v_k are i.i.d. and mutually independent Gaussian noises. $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, and $C \in \mathbb{R}^{p\times n}$ are the system matrix, input matrix, and measurement matrix, respectively. As is customary in the field of LQG control, we make the following assumptions throughout the paper: the pair (A, B) is stabilizable, the pair (A, C) is detectable, and $(A, \sqrt{Q})$ is stabilizable.

The LQG control objective is to design a feedback control law to minimize the cost

$$J = \mathbb{E}\Big(\frac{1}{2}x_N^{\mathrm{T}} M x_N + \frac{1}{2}\sum_{k=k_0}^{N-1}(x_k^{\mathrm{T}} W x_k + u_k^{\mathrm{T}} U u_k)\Big), \tag{2}$$

where $M \in \mathbb{R}^{n\times n}$, $W \in \mathbb{R}^{n\times n}$ and $U \in \mathbb{R}^{m\times m}$ are positive definite matrices. To solve such a problem, an estimator is usually deployed, e.g., a Kalman filter of the form

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_k + Bu_k \\ P_{k+1|k} &= AP_kA^{\mathrm{T}} + Q \\ K_{k+1} &= P_{k+1|k}C^{\mathrm{T}}(CP_{k+1|k}C^{\mathrm{T}} + R)^{-1} \\ \hat{x}_{k+1} &= \hat{x}_{k+1|k} + K_{k+1}(y_{k+1} - C\hat{x}_{k+1|k}) \\ P_{k+1} &= P_{k+1|k} - K_{k+1}CP_{k+1|k}\,, \end{aligned} \tag{3}$$

where $\hat{x}_k$ is the estimated state, P_k is the error covariance, and K_k is the gain matrix. $\hat{x}_{k+1|k}$ and $P_{k+1|k}$ are the predictions of $\hat{x}_k$ and P_k, respectively. With the above Kalman filter, the LQG control law can be given by

$$u_k = L_k\hat{x}_k \tag{4}$$

where $L_k = -(B^{\mathrm{T}}S_{k+1}B + U)^{-1}B^{\mathrm{T}}S_{k+1}A$, with S_{k+1} satisfying

$$S_k = A^{\mathrm{T}}S_{k+1}A + W - A^{\mathrm{T}}S_{k+1}B(B^{\mathrm{T}}S_{k+1}B + U)^{-1}B^{\mathrm{T}}S_{k+1}A\,. \tag{5}$$

According to [18], when $N \to \infty$, the estimation error covariance P_k converges as $k \to \infty$ to a steady-state solution P, satisfying the Riccati equation

$$P = APA^{\mathrm{T}} + Q - APC^{\mathrm{T}}(CPC^{\mathrm{T}} + R)^{-1}CPA^{\mathrm{T}}, \tag{6}$$

yielding the gain matrix K_k in steady state as

$$K = PC^{\mathrm{T}}(CPC^{\mathrm{T}} + R)^{-1}. \tag{7}$$

In addition, according to [19], when $N \to \infty$, the solution S_k converges as $k \to \infty$ to a steady-state solution S, satisfying the Riccati equation

$$S = A^{\mathrm{T}}SA + W - A^{\mathrm{T}}SB(B^{\mathrm{T}}SB + U)^{-1}B^{\mathrm{T}}SA, \tag{8}$$

yielding the control gain L_k in steady state as

$$L = -(B^{\mathrm{T}}SB + U)^{-1}B^{\mathrm{T}}SA. \tag{9}$$

B. Replay Attack

In Fig. 1, we can see that during the transmission of data from the sensor to the estimator over the network, there is a possibility of an adversary attacking the transmission channel and maliciously tampering with the data. Specifically, we consider an adversary launching a replay attack in this paper, which has access to the following resources [9], [10]:

(i) The adversary can monitor and record the true sensor outputs y_k for all k.

(ii) The adversary can modify both the control signals u_k and sensor signals y_k to arbitrary values.

Without loss of generality, we denote the time when the adversary launches the attack as time 0. The considered replay attack strategy is given as follows.

1) From time $-T$ to time -1, the attacker records a sequence of sensor measurements y_k.
2) From time 0 to time $T-1$, the attacker injects designed u_k^a to control signal u_k to damage the physical plant. Meanwhile, the attacker replays the recorded sensor measurements to tamper the true sensor outputs, i.e.,

$$y_k^a = y_{k-T}, \quad 0 \le k \le T-1. \tag{10}$$

Here we assume that the control system has run for a long time (already achieved a steady state) before it is attacked, i.e., the LQG control system has already reached its steady state before time $-T$.

According to (7)-(9), the steady-state Kalman filter and LQG controller are given by

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_k + Bu_k \\ \hat{x}_k &= \hat{x}_{k|k-1} + Kz_k \end{aligned} \tag{11}$$

with $z_k = y_k - C\hat{x}_{k|k-1}$ and $u_k = L\hat{x}_k$. Thus, under attack ($0 \le k \le T-1$), the physical plant takes the form

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k^a + w_k \\ y_k &= Cx_k + v_k. \end{aligned} \tag{12}$$

The Kalman filter and LQG controller take the form

$$\begin{aligned} \hat{x}_{k+1|k} &= A\hat{x}_k + Bu_k \\ \hat{x}_k &= \hat{x}_{k|k-1} + Kz_k' \end{aligned} \tag{13}$$

with $z_k' = y_k^a - C\hat{x}_{k|k-1}$ and $u_k = L\hat{x}_k$.

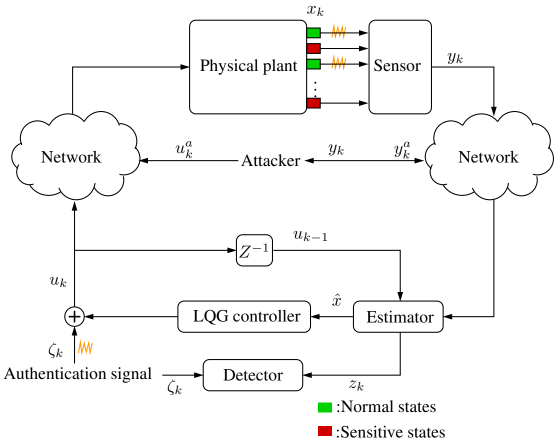

Fig. 1. The framework of an LQG control system.

C. Replay Attack Detection

To detect abnormal data from the received sensor measurements, at the controller side a failure detector $g(z_k)$ needs to be deployed. See Fig. 1 for the framework of the considered LQG control problem with the replay attack and the failure detector. A widely used detector is termed as χ^2 detector [20], [21], i.e.,

$$g(z_k) = z_k^{\mathrm{T}}\mathcal{P}^{-1}z_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \nu \tag{14}$$

where $\mathcal{P} = CPC^{\mathrm{T}} + R$, $\mathcal{H}_0$ means that no alarm is triggered, $\mathcal{H}_1$ means that the detector alarms, and ν is a threshold.

For such a χ^2 detector, however, it is worth noting that the replay attack may still be successfully launched, as shown in [6, Theorem 2]. In view of this, to detect replay attacks, the watermarking mechanism is proposed in [6], where an i.i.d. Gaussian authentication signal is added to the control inputs, i.e.,

$$u_k = L\hat{x}_k + \zeta_k \tag{15}$$

where $\zeta_k \sim \mathcal{N}(0, \Lambda)$ is the i.i.d. Gaussian authentication signal. The resulting control performance and detection effect are given below.

Lemma 1: [9, Theorem 5] The LQG performance loss caused by adding ζ_k is

$$\Delta J = \mathrm{Tr}[(U + B^{\mathrm{T}}SB)\Lambda]. \tag{16}$$

Lemma 2: [9, Theorem 6] In the absence of an attack,

$$\mathbb{E}[z_k^{\mathrm{T}}\mathcal{P}z_k] = p. \tag{17}$$

Under attack

$$\lim_{k\to\infty} \mathbb{E}[z_k'^{\mathrm{T}}\mathcal{P}z_k'] = p + 2\mathrm{Tr}(\mathrm{C}^{\mathrm{T}}\mathcal{P}^{-1}\mathrm{C}\mathcal{U}) \tag{18}$$

where $\mathcal{U} := \sum_{j=0}^{\infty} \mathcal{A}^j B\Lambda B^{\mathrm{T}}(\mathcal{A}^j)^{\mathrm{T}}$ with $\mathcal{A} = (A+BL)(I-KC)$. In addition, $\mathcal{U}$ can be solved by equation

$$\mathcal{U} - B\Lambda B^{\mathrm{T}} = \mathcal{A}\mathcal{U}\mathcal{A}^{\mathrm{T}}. \tag{19}$$

Here we discuss the case that $\mathcal{A}$ is stable ($\rho(\mathcal{A}) < 1$), otherwise $\lim_{k\to\infty}\mathbb{E}[z_k'\mathcal{P}z_k'^{\mathrm{T}}] = \infty$ and the replay attack can and must be detected.

By combining the above two results, a constrained optimization problem can be formulated for the design of the covariance Λ of authentication signal, in the sense of maximizing the detection effect, i.e., $2\mathrm{Tr}(\mathrm{C}^{\mathrm{T}}\mathcal{P}^{-1}\mathrm{C}\mathcal{U})$, while limiting the control performance loss, i.e., ΔJ.

D. Problem Statement

When the system is functioning normally (without attack), the authentication signal is essentially a disturbance input that affects the control performance of the physical plant. However, in many practical applications, some plant states are highly sensitive to disturbances and can only tolerate limited amounts of disturbance or none at all. For instance, maintaining a stable temperature in chemical processes is crucial for certain reactions [14], [15]. Similarly, in robotics, lateral movements of a robot must be carefully controlled to maintain accurate tracking [16], while the front joint of a manipulator must avoid disturbances to ensure precise control of the end-point [17]. To sum up, sensitive states are states that are expected to be limitedly affected by authentication signals. More formally, the sensitive states are measured in the sense of mean square error, i.e., a state $x_{i,k}$ is sensitive if it should satisfy

$$\mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2] \leq \eta_i \tag{20}$$

where η_i is the tolerance threshold of the sensitive state $x_{i,k}$, $\bar{x}_{i,k}$ is the plant state impacted by the authentication signal.

The previously mentioned approach does not take into account the impact of the authentication signal on sensitive states. As a result, these authentication signals may significantly impact sensitive states, potentially violating the condition described in Equation (20). Motivated by this issue, this paper aims to propose a new authentication signal design method that can satisfy Equ. (20). Specifically, this paper aims to address the following two problems:

1) What conditions must be met to ensure that the authentication signal has a limited effect on sensitive states (i.e., Equation (20) holds), and even has no effect on them (i.e., Equation (20) holds with $\eta_i = 0$)?
2) How can the authentication signal be optimally designed to satisfy these conditions?

Remark 1: It is assumed that the length of the recorded sensor measurements T is long enough. In practice, if the attacker can only record a short segment of sensor measurements, he/she can loop the recorded data to form long replay data.

Remark 2: The attack signal u_k^a can be arbitrarily chosen because the detector cannot receive true sensor measurements under the replay attack.

Remark 3: One possible approach to reducing the impact of the authentication signal on sensitive states is to increase the weight W in Equation (2) for the sensitive states. However, this approach has some drawbacks. Firstly, the weight W is usually designed to balance the trade-off between the tracking performance and control cost, and adjusting it to design the authentication signal may not be reasonable or optimal. Secondly, even if it were permissible to adjust W, it is challenging to determine an appropriate value for W since it depends on the specific system and the importance of the sensitive states. Lastly, even if W were set to a large value, the impact of the authentication signal on sensitive states cannot be completely eliminated since it is not practical to set the weight to infinity.

III. Authentication Signal Design

In this section, we will present the conditions under which the authentication signal has a limited effect or no effect on sensitive states. Then, optimization problems are formulated for designing the authentication signal.

A. Effect Analysis of Authentication Signal on Sensitive States

This subsection will provide the conditions under which the authentication signal has a limited effect and even no effect on sensitive states. We will focus on the scenario where the system is not under attack. This is because, during an attack, the attacker can choose u_a to dominate the system performance, and our main objective is to detect the attack.

The following proposition presents the condition that the authentication signal has a limited effect on sensitive states.

Proposition 1: The mean square errors are limited to $\mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2] \le \eta_i$ for all $i \in \mathcal{F}$ if and only if

$$e_i \Upsilon e_i^{\mathrm{T}} \le \eta_i, \quad \forall i \in \mathcal{F} \tag{21}$$

where $e_i \in \mathbb{R}^{1\times n}$ is a row vector with all elements being zero except the i^{th} element as 1, Υ is the solution of the equation $\Upsilon - (A+BL)\Upsilon(A+BL)^{\mathrm{T}} = B\Lambda B^{\mathrm{T}}$, $\mathcal{F} := \{i | x_{i,k} \text{ is sensitive}\}$, and $x_{i,k}$ is the i^{th} element of x_k.

Proof: Without the authentication signal, system (1) with LQG controller is

$$\begin{aligned} x_{k+1} &= (A+BL)x_k + BL\tilde{x}_k + w_k \\ y_k &= Cx_k + v_k \end{aligned} \tag{22}$$

where $\tilde{x}_k := \hat{x}_k - x_k$, and the iteration of $\tilde{x}_k$ can be obtained according to (1) and (11), i.e.,

$$\begin{aligned} \tilde{x}_{k+1} &= A\tilde{x}_k + K(y_{k+1} - C\hat{x}_{k+1|k}) - w_k \\ &= (A-KCA)\tilde{x}_k + Cw_k - w_k + v_{k+1}. \end{aligned} \tag{23}$$

Since we have assumed that the LQG system has run for a long time before it is attacked, we denote the time instant when the system starts to work as $-\infty$. According to (22), the sensitive states can be solved as

$$x_{i,k} = e_i \sum_{j=-\infty}^{k-1} (A+BL)^{k-j-1}(BL\tilde{x}_j + w_j), \ i \in \mathcal{F}. \tag{24}$$

With the authentication signal, the controller becomes (15), and the system (1) becomes

$$\begin{aligned} \bar{x}_{k+1} &= (A+BL)\bar{x}_k + B\zeta_k + BL\bar{\tilde{x}}_k + w_k \\ \bar{y}_k &= C\bar{x}_k + v_k \end{aligned} \tag{25}$$

where $\bar{x}_k$ denotes the system state with the authentication signal added to the system. Let $\bar{\tilde{x}}_k := \bar{\hat{x}}_k - \bar{x}_k$. According to the authentication signal affected version of (1) and (11), the iteration of $\bar{\tilde{x}}_k$ can be obtained as

$$\begin{aligned} \bar{\tilde{x}}_{k+1} &= A\bar{\tilde{x}}_k + K(\bar{y}_{k+1} - C\bar{\hat{x}}_{k+1|k}) - w_k \\ &= (A-KCA)\bar{\tilde{x}}_k + Cw_k - w_k + v_{k+1}. \end{aligned} \tag{26}$$

where $\bar{\hat{x}}_{k+1|k}$ is the prediction of the estimator. Comparing (23) with (26), we have $\bar{\tilde{x}}_k = \tilde{x}_k, \forall k$.

According to (24), (25) and the fact that $\bar{\tilde{x}}_k = \tilde{x}_k$, for $i \in \mathcal{F}$, the sensitive state impacted by the authentication signal can be solved as

$$\begin{aligned} \bar{x}_{i,k} &= e_i \sum_{j=-\infty}^{k-1} (A+BL)^{k-j-1}(B\zeta_j + BL\bar{\tilde{x}}_j + w_j) \\ &= x_{i,k} + e_i \sum_{j=-\infty}^{k-1} (A+BL)^{k-j-1}B\zeta_j. \end{aligned} \tag{27}$$

According to (27), there have

$$\begin{aligned} &\mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2)] = \mathbb{E}\Big[(e_i \sum_{j=-\infty}^{k-1} (A+BL)^{k-j-1}B\zeta_j)^2\Big] \\ &= \sum_{j=-\infty}^{k-1} e_i(A+BL)^{k-j-1}B\Lambda B^{\mathrm{T}}((A+BL)^{k-j-1})^{\mathrm{T}}e_i^{\mathrm{T}}. \end{aligned} \tag{28}$$

Let $\mathfrak{F}_i(k) := \mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2], \forall i \in \mathcal{F}$. It is clear that $\mathfrak{F}_i(k)$ is monotonically increasing as k increase. So $\mathbb{E}((\bar{x}_{i,k} - x_{i,k})^2) \le \eta_i, \forall k, \forall i \in \mathcal{F}$ if and only if $\lim_{k\to\infty} \mathfrak{F}_i(k) \le \eta_i, \forall i \in \mathcal{F}$, i.e.,

$$e_i \Upsilon e_i^{\mathrm{T}} \le \eta_i, \forall i \in \mathcal{F} \tag{29}$$

where $\Upsilon := \sum_{j=0}^{\infty}((A+BL)^j B\Lambda B^{\mathrm{T}}((A+BL)^j)^{\mathrm{T}}$. Note that it can be easily verified that such Υ is the solution of the equation $\Upsilon - B\Lambda B^{\mathrm{T}} = (A+BL)\Upsilon(A+BL)^{\mathrm{T}}$. Then the proof is complete. ■

A particular scenario of Proposition 1 arises when $\eta_i = 0$, indicating that the designer has a preference of no impact on the sensitive states, under which the condition in Proposition 1 can be simplified.

Proposition 2: The authentication signal ζ_k does not affect the sensitive states of system (25), i.e., $\mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2] = 0$, if and only if

$$\Omega B \Pi = 0. \tag{30}$$

where $\Pi\Pi^{\mathrm{T}} = \Lambda$,

$$\Omega := \begin{bmatrix} E \\ E(A+BL) \\ \vdots \\ E(A+BL)^{n-1} \end{bmatrix}. \tag{31}$$

with $E := [e_{i_1}^{\mathrm{T}}, e_{i_2}^{\mathrm{T}}, ..., e_{i_q}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{q\times n}$, where $i_1, i_2, ..., i_q \in \mathcal{F}$ and $q = |\mathcal{F}|$.

Proof: According to (28), $\mathbb{E}[(\bar{x}_{i,k} - x_{i,k})^2] = 0$ is equivalent to $\forall k, \forall i \in \mathcal{F}$,

$$\sum_{j=-\infty}^{k-1} e_i(A+BL)^{k-j-1}B\Pi\Pi^{\mathrm{T}}((A+BL)^{k-j-1}B)^{\mathrm{T}}e_i^{\mathrm{T}} = 0. \tag{32}$$

Note that the above equation is equivalent to

$$e_i(A+BL)^s B\Pi = 0\,, \quad \forall i \in \mathcal{F}, \forall s \in \mathbb{N}. \tag{33}$$

By Cayley-Hamilton theorem [22], $(A+BL)^n$ is a linear combination of $(A+BL)^s$ for $s = 0, 1, ..., n-1$. It is clear that (33) is equivalent to saying

$$e_i(A+BL)^s B\Pi = 0\,, \quad \forall i \in \mathcal{F}, \forall s \in \{0, 1, \dots, n-1\}.$$

This thus completes the proof by calling the definition of E. ■

B. Optimal Authentication Signal Design

To maximize the detection effect and limit both the performance loss and effect to sensitive states, according to Lemma 1, Lemma 2 and Proposition 1, an optimal problem is constructed as

$$\begin{aligned} \max_{\Lambda \in \mathbb{P}^+} \quad & 2\mathrm{Tr}(\mathrm{C}^{\mathrm{T}}\mathcal{P}^{-1}\mathrm{C}\mathcal{U}) \\ \text{s.t.} \quad & \mathrm{Tr}[(U + B^{\mathrm{T}}SB)\Lambda] \le \delta \\ & e_i \Upsilon e_i^{\mathrm{T}} \le \eta_i,\ i \in \mathcal{F} \end{aligned} \tag{34}$$

where

$$\mathcal{U} - \mathcal{A}\mathcal{U}\mathcal{A}^{\mathrm{T}} = B\Lambda B^{\mathrm{T}} \tag{35}$$

and

$$\Upsilon - (A+BL)\Upsilon(A+BL)^{\mathrm{T}} = B\Lambda B^{\mathrm{T}}. \tag{36}$$

According to the definition of $\mathcal{U}$ and Υ, they are linear to Λ. Thus functions $2\mathrm{Tr}(\mathrm{C}^{\mathrm{T}}\mathcal{P}^{-1}\mathrm{C}\mathcal{U})$, $\mathrm{Tr}[(U + B^{\mathrm{T}}SB)\Lambda]$, and $e_i\Upsilon e_i^{\mathrm{T}},\ i \in \mathcal{F}$ are all linear to Λ. So the feasible set of problem (34) is convex since it is the intersection of a polyhedron and a positive semi-definite cone ($\Lambda \in \mathbb{P}^+$). Given that the objective function is convex, we conclude that the problem (34) is a convex optimization problem, specifically a semi-definite programming (SDP) problem.

In addition, to make it more convenient for solving (34) with optimization toolboxes, Eqns. (35) and (36) can be solved by vectorization operation [23, Proposition 10.4]. For example, $\mathcal{U}$ can be solved as

$$\mathcal{U} = \sum_{j=1}^{n} (\mathcal{E}_j^n)^{\mathrm{T}}\Big\{[I_{n^2} - (\mathcal{A}\otimes\mathcal{A})]^{-1}(B\otimes B)\sum_{i=1}^{n}\mathcal{E}_i^m \Lambda e_i^m\Big\}(e_j^n)^{\mathrm{T}}. \tag{37}$$

where $\mathcal{E}_j^n \in \mathbb{R}^{n^2\times n}$ are column-wise block matrices consisting of n blocks of size $n \times n$, with an identity matrix only in the i^{th} block and the others are all zeros, and $\mathcal{E}_i^m \in \mathbb{R}^{m^2\times n}$ share the same structure with $\mathcal{E}_j^n$. $e_j^n \in \mathbb{R}^n$ are column vectors, whose i^{th} element is 1 and the others are all zeros, and $e_i^m \in \mathbb{R}^m$ share the same structure with e_j^n.

A special case of (34) is $\eta_i = 0, \forall i \in \mathcal{F}$. According to Proposition 2, in this case, the second constraint in problem (34) becomes $\Omega B\Lambda = 0$.

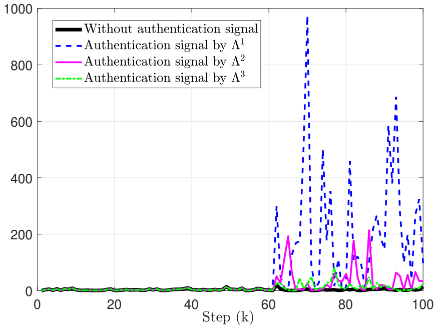

Fig. 2. The output of the detector with different authentication signal.

IV. Simulation Verification

In this section, simulation results are given. The physical plant model is a chemical reactor borrowed from [24]. The linearized model of the chemical reactor is

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} &= \begin{bmatrix} -0.0010 & -0.0254 & 0 & -0.2052 \\ 0 & -0.0036 & 0.004 & 0.0048 \\ 0 & -0.4571 & -0.0429 & 0 \\ 0 & 0 & 0 & -0.05 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0.78 \\ 0 & 0 \\ 40 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ y &= \begin{bmatrix} x_1 & x_2 \end{bmatrix}^{\mathrm{T}} \end{aligned} \tag{38}$$

where x_1, x_2, x_3 and x_4 are the concentration of reactant, the temperature of the reactor, the temperature of the jacket and the liquid volume of the reactor, respectively. In addition, to obtain the discrete-time model, the continuous-time model (38) is discretized with time step 0.1s. The noises added to the discrete model are set as $Q = R = 0.01I_4$.

The LQG controller is designed with parameters $W = I_4$ and $U = I_2$. The performance loss limitation δ is set as 4. For comparison, if the sensitive states are not considered, i.e., the constrains $e_i\Upsilon e_i^{\mathrm{T}} \le \eta_i,\ i \in \mathcal{F}$ are removed in (34), the solution is

$$\Lambda^1 = \begin{bmatrix} 0.1181 & 0 \\ 0 & 0 \end{bmatrix}. \tag{39}$$

Suppose the temperature of the reactor (x_2) and the temperature of the jacket (x_3) are the sensitive states. Setting $\eta_2 = 0.0001$ and $\eta_3 = 0.5$, the solution to (34) is

$$\Lambda^2 = \begin{bmatrix} 0.0304 & 0.7463 \\ 0.7463 & 18.3445 \end{bmatrix}. \tag{40}$$

In addition, if the sensitive state does not allow any additional noise, i.e., $\eta_2 = \eta_3 = 0$, the solution is

$$\Lambda^3 = \begin{bmatrix} 0 & 0 \\ 0 & 24.6921 \end{bmatrix}. \tag{41}$$

Then we add the designed authentication signal into the system model, and the replay attacks begin at step 60. Simulation results are presented in Figs. 2-4. From Fig. 2,

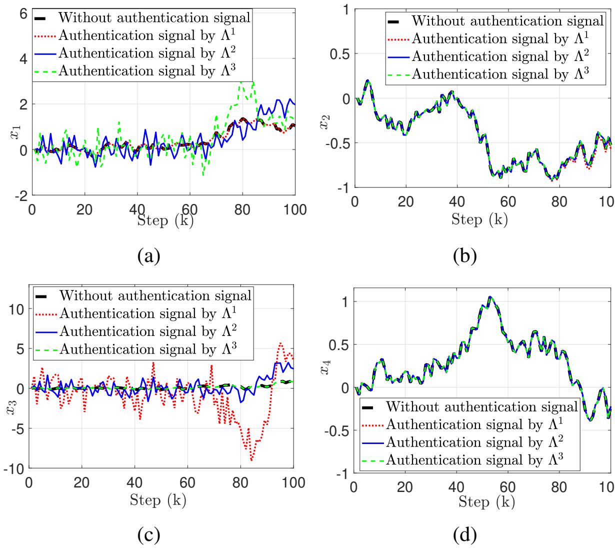

Fig. 3. The state of the system with different authentication signal.

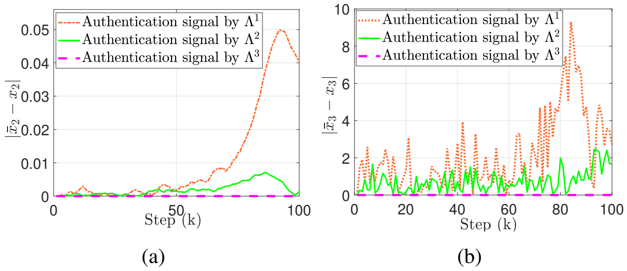

Fig. 4. Effect of different authentication signal to sensitive states.

we see that the amplitude of $g(z_k)$s' increases in different degrees. In detail, the authentication signal generated by Λ^1 is the most obvious, followed by Λ^2, and Λ^3 is the least. From Figs. 3 and 4, with the authentication signal generated by Λ^1, the sensitive states suffer from a great amount of noise, while with our method, a smaller amount of noise (Λ^2) or no noise (Λ^3) would affect sensitive states, which verifies the effectiveness our proposed method.

To conclude, from Figs. 2, 3 and 4, we find that although our method would sacrifice detection performance, it can effectively limit the effect of authentication signal to sensitive states.

V. Conclusion

This paper presented a method for designing an authentication signal that limits its impact on the sensitive states of a physical plant. We first measured the effect of authentication to sensitive states by mean square error. Using geometric control methods, we derived conditions under which sensitive states would be subject to limited effect or even be free of the authentication signal. We then constructed optimization problems containing with these conditions and utilized optimization tools to find the optimal authentication signal. Finally, we verified the effectiveness of the proposed authentication signal by comparing simulation results.

References

[1] D. Zhang, Q.-G. Wang, G. Feng, Y. Shi, and A. V. Vasilakos, "A survey on attack detection, estimation and control of industrial cyber¨cphysical systems," *ISA Transactions*, vol. 116, pp. 1–16, 2021.

[2] M. Zhang, S. Dong, P. Shi, G. Chen, and X. Guan, "Distributed observer-based event-triggered load frequency control of multiarea power systems under cyber attacks," *IEEE Transactions on Automation Science and Engineering*, 2022.

[3] J. P. Farwell and R. Rohozinski, "Stuxnet and the future of cyber war," *Survival*, vol. 53, no. 1, pp. 23–40, 2011.

[4] J. Slay and M. Miller, "Lessons learned from the maroochy water breach," in *Critical Infrastructure Protection; IFIP International Federation for Information Processing; 253; Springer Series in Computer Science*, 2008.

[5] M. Zhang, Z. Wu, J. Yan, R. Lu, and X. Guan, "Attack-resilient optimal pmu placement via reinforcement learning guided tree search in smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1919–1929, 2022.

[6] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 911–918, 2009.

[7] A. Khazraei, H. Kebriaei, and F. R. Salmasi, "A new watermarking approach for replay attack detection in lqg systems," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5143–5148, IEEE, 2017.

[8] H. S. Sanchez, D. Rotondo, T. Escobet, V. Puig, J. Saludes, and J. Quevedo, "Detection of replay attacks in cyber-physical systems using a frequency-based signature," *Journal of the Franklin Institute*, vol. 356, no. 5, pp. 2798–2824, 2019.

[9] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2013.

[10] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[11] C. Fang, Y. Qi, P. Cheng, and W. X. Zheng, "Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems," *Automatica*, vol. 112, p. 108698, 2020.

[12] B. Tang, L. D. Alvergue, and G. Gu, "Secure networked control systems against replay attacks without injecting authentication noise," in *2015 American Control Conference (ACC)*, pp. 6028–6033, IEEE, 2015.

[13] D. Ye, T. Zhang, and G. Guo, "Stochastic coding detection scheme in cyber-physical systems against replay attack," *Information Sciences*, vol. 481, pp. 432–444, 2019.

[14] T. Urit, M. Li, T. Bley, and C. Löser, "Growth of kluyveromyces marxianus and formation of ethyl acetate depending on temperature," *Applied microbiology and biotechnology*, vol. 97, no. 24, pp. 10359–10371, 2013.

[15] Z. Sun, Y. Tang, T. Iwanaga, T. Sho, and K. Kida, "Production of fuel ethanol from bamboo by concentrated sulfuric acid hydrolysis followed by continuous ethanol fermentation," *Bioresource Technology*, vol. 102, no. 23, pp. 10929–10935, 2011.

[16] R. Rajamani, *Vehicle dynamics and control*. Springer Science & Business Media, 2011.

[17] F. L. Lewis, D. M. Dawson, and C. T. Abdallah, *Robot manipulator control: theory and practice*. CRC Press, 2003.

[18] B. D. Anderson and J. B. Moore, *Optimal filtering*. Courier Corporation, 2012.

[19] D. P. Bertsekas, "Dynamic programming and optimal control," *Athena Scientific*,, 1995.

[20] J. Antoch, "A guide to chi-squared testing," *Computational Statistics & Data Analysis*, vol. 23, no. 4, pp. 565–566, 1997.

[21] R. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, 1971.

[22] C. D. Meyer, *Matrix analysis and applied linear algebra*, vol. 71. Siam, 2000.

[23] J. D. Hamilton, *Time series analysis*. Princeton university press, 1994.

[24] J. Corriou, "Chapter 10 - optimal control," in *Coulson and Richardson's Chemical Engineering (Fourth Edition)* (S. Rohani, ed.), pp. 403–440, Butterworth-Heinemann, 2017.