A Novel Stochastic Event-Based Feedback Watermark for Replay Attack Detection

Xudong Zhao(✉), Xinyu Wang, Fengde Xu, and Yuhan Wang

Key Laboratory of Intelligent Control and Optimization for Industrial Equipment, Dalian University of Technology, Dalian 116024, China
xdzhaohit@gmail.com

Abstract. The security of cyber physical systems (CPSs) is the premise of resource sharing in modern industry, which has attracted considerable attention of researchers. Many effective methods have been proposed to defend CPSs, among which physical watermark is one prevailing approach which can enhance the replay attack detection capability of CPSs. In order to make the watermarks work more effective, a novel stochastic event-based feedback physical watermark is proposed in this paper to detect replay attacks. We formulate first the problem taking into account the Kalman filter, the linear quadratic Gaussian (LQG) optimal controller and the χ^2 detector. Then, we characterize the LQG performance loss and the probability of adding a physical watermark in two different scenarios: the system operates with and without replay attacks. It is proved that the probability of adding a watermark signal will increase when replay attacks exist. Furthermore, we discuss the performance of the χ^2 detector under the framework of our approach. Finally, numerical simulations are verified the theoretical results.

Keywords: CPSs security · Stochastic event-based feedback physical watermark · Replay attacks

1 Introduction

Cyber physical systems integrate distributed networks of smart sensors and computational elements with physical plants, which achieve better efficiency and productivity than traditional control systems [1]. However, the wide usage of communication networks may bring many drawbacks in security, such as the Stuxnet Worm on Iranian centrifuges and the attack on Venezuelan hydropower stations, leading to a serious depression of the economy or even threaten the national security. Therefore, it is in an urgent need to model and analyze the CPSs security.

This work is supported by the National Natural Science Foundation of China (U21A20477, 61722302, 61573069, 62203064), the Fundamental Research Funds for the Central Universities (DUT19ZD218, DUT22ZD402), the Liaoning Revitalization Talents Program under Grant XLYC1907140, and National Major Science and Technology Project (J2019-V-0010-0105).

H. R. Karimi and N. Wang (Eds.): S-Cube 2022, LNICST 487, pp. 107–119, 2023.
https://doi.org/10.1007/978-3-031-34899-0_7

Unsurprisingly, the CPSs security has been extensively studied in recent years. The work [2] has considered two main kinds of attacks, i.e., injection attacks and denial of service (DoS) attacks. Compared with DoS attacks, the injection attacks require comprehensive information about the system, making them more difficult to be detected. Therefore, in the following, we concentrate on injection attacks, where the attackers alter sensor measurements and affect date integrity.

Replay attacks are the main classes of injection attacks, which need to replay the previous sensor measurements for the purpose of deceiving the CPSs. It should be pointed out that replay attacks have the ability to bypass passive detection methods in regardless of the knowledge of the system. However, adding physical watermarks is a particular active detection method, which is helpful to defend against replay attacks first provided in [3]. According to the work [3], the authors in work [4] have extended the results by presenting a more general watermarks scheme, and proposed a Cross-Correlator to detect replay attacks. In addition, Mo et al. [5] also have designed the correlated physical watermarks using stationary Gaussian processes. Weerakkody et al. [6] have focused on the physical watermarks in the presence of data packet dropouts in the system. In work [7], the authors have been concerned with finding an algorithm that can generate watermarks to detect replay attacks in the fact of unknown system parameters. However, the above physical watermarks on the one hand can detect replay attacks easily, but, on the other hand, they would like to degrade the control performance to a certain degree. In order to reduce the loss of control performance, Fang et al. [8] have been interesting in obtaining a periodic schedule to add watermarks. Besides, the authors [9] have paid attention to a multiplicative sensor watermarks. Furthermore, Miao et al. [10,11] have developed a sub-optimal scheme to add watermark signals over a finite time horizon by resorting to the game theoretic approach.

It is obvious that most of the physical watermarks are closely relevant to the time evolutions, while few work focuses on event-based physical watermarks. Also, event-based method is a hot spot and widely used to reduce the communication cost [12,13]. When the predefined condition is met, the sensor measurements are immediately transmitted to the remote estimator over the wireless communication network. Therefore, we find that the event scheduler can utilize real-time data and provide more information for decision whether to add a watermark signal. Motivated by the above, we provide in the paper an event-based feedback watermark, which, on the one hand, improve the detection rate subject to replay attacks, and on the other hand, ensure system performance.

In this paper, a novel stochastic event-based feedback physical watermark is proposed for detecting replay attacks. The system performances are studied after adding the physical watermarks. In addition, our watermarks are proved to effectively improve detection rates of χ^2 detector. The main contributions of this paper can be divided into the following:

1. We propose a novel stochastic event-based feedback physical watermark, and prove that the probability of adding a watermark is a constant without replay attacks. In such scenario, we also investigate the system performance loss.
2. We provide the boundary of the probability of launching a watermark under replay attacks. We show that the probability of launching a watermark increases in the case of replay attacks exist.
3. We show the effectiveness of stochastic event-based feedback physical watermarks for detecting replay attacks.

The rest of this paper is organized as follows: Sect. 2 setup the problem taking into account the Kalman filter, the LQG controller and the χ^2 detector. Section 3 proposes a stochastic event-based feedback watermark and characterizes the system performances without replay attacks. Section 4 shows the validity of our proposed watermarks under replay attacks. Section 5 gives some numerical simulations to show the effectiveness of our watermarks, followed by the conclusion in Sect. 6.

2 Problem Setup

We introduce a linear time invariant (LTI) system:

$$x_{k+1} = Ax_k + Bu_k + w_k, \tag{1}$$
$$y_k = Cx_k + v_k \tag{2}$$

where x_k, u_k and y_k are the system state, control input and the sensor measurement, respectively. The initial state $x_0 \sim \mathcal{N}(0, \Sigma)$, while w_k and v_k are mutually independent, w_k and v_k are independent identically distributed (i.i.d.) Gaussian variables with covariance Q and R, respectively. Furthermore, we assume that (A, C) is detectable and $\left(A, Q^{\frac{1}{2}}\right)$ is stabilizable.

A Kalman filter is used to provide optimal state estimate of state x_k:

$$\hat{x}_{0|-1} = 0, P_{0|-1} = \Sigma, \tag{3}$$
$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k, \tag{4}$$
$$P_{k+1|k} = AP_{k|k}A^T + Q, \tag{5}$$
$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}, \tag{6}$$
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}), \tag{7}$$
$$P_{k|k} = P_{k|k-1} - K_kCP_{k|k-1}, \tag{8}$$

It should be pointed out that the gain K_k will converge exponentially if (A, C) is detectable [14]. Hence, the filter can be expressed as a fixed gain estimator:

$$P \triangleq \lim_{k\to\infty} P_{k|k-1}, K \triangleq PC^T(CPC^T + R)^{-1}, \tag{9}$$
$$\hat{x}_{k+1|k} = A\hat{x}_{k|k} + Bu_k, \tag{10}$$
$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K(y_k - C\hat{x}_{k|k-1}). \tag{11}$$

Using $\hat{x}_{k|k}$ generated by the Kalman filter, the controller wants to minimize LQG cost:

$$J = \lim_{M\to\infty} \mathbb{E}\{\frac{1}{M}[\sum_{k=0}^{M-1}(x_k^T W x_k + u_k^T U u_k)]\}, \tag{12}$$

where $W \succ 0$ and $U \succ 0$. The optimal control input adopts a steady-state strategy:

$$u_k^* = -(B^T S B + U)^{-1} B^T S A \hat{x}_{k|k} = L\hat{x}_{k|k}, \tag{13}$$

where S is the unique positive definite matrix satisfies the following Riccati equation:

$$S = A^T S A + W - A^T S B (B^T S B + U)^{-1} B^T S A. \tag{14}$$

The cost function in this case becomes a constant:

$$J = trace(SQ) + trace[(A^T S A + W - S)(P - KCP)]. \tag{15}$$

2.1 χ^2 Detector

χ^2 detector is very commonly used for anomaly detection in the system [14], which utilizes the statistical properties of the Kalman filter innovation:

Lemma 1. *The innovation $z_i = y_i - C\hat{x}_{i|i-1} \sim \mathcal{N}(0, \mathscr{P})$ is a i.i.d. Gaussian random variable where $\mathscr{P} = CPC^T + R$ [15].*

Let

$$g_k = \sum_{i=k-\mathscr{K}+1}^{k} z_i^T \mathscr{P}^{-1} z_i \underset{\mathscr{H}_1}{\overset{\mathscr{H}_0}{\lessgtr}} \eta, \tag{16}$$

where $\mathscr{K}$ is the window size and η is the threshold. $\mathscr{H}_1$ denotes replay attack exists and $\mathscr{H}_0$ is on the contrary. From Lemma 1, it is trivial to show that g_k has a χ^2 distribution with $m\mathscr{T}$ degrees of freedom when the system is operating normally.

Define the false alarm and detection rate as α_k and β_k, respectively:

$$\alpha_k \triangleq Pr(g_k > \eta | \mathscr{H}_0), \beta_k \triangleq Pr(g_k > \eta | \mathscr{H}_1). \tag{17}$$

2.2 Replay Attack and Physical Watermark

In this paper, we concentrate on replay attack. The malicious entities can record all sensor measurements and arbitrarily modify them into $y_k's$. We have the following results from work [4]:

Lemma 2. *If $\mathscr{A} \triangleq (A + BL)(I - KC)$ is stable*

$$\lim_{k\to\infty} \beta_k = \alpha_k. \tag{18}$$

On the contrary, if it is unstable

$$\lim_{k\to\infty} \beta_k = 1 \tag{19}$$

This lemma shows that an attacker can fool the χ^2 detector if and only if $\mathscr{A}$ is stable. Therefore, we assume that $\mathscr{A}$ is stable in the rest of this article. In addition, physical watermarks are added to defend against such attacks,

$$u_k = u_k^* + \gamma_k \Delta u_k, \tag{20}$$

where $\gamma_k = 1$ means adding a physical watermark at time k and $\gamma_k = 0$ is the opposite, $\Delta u_k \sim \mathcal{N}(0, \mathscr{Q})$ is the watermark that follows the i.i.d. Gaussian distribution, and for all k, it is independent of u_k^*, w_k and v_k. It is worth noting that in the existing approaches such as [3,4], the watermark is always added, i.e., $\gamma_k \equiv 1$. However, the LQG cost of this approach is too high. In the next section, we will use a stochastic event-based approach to design γ_k.

3 Stochastic Event-Based Feedback Physical Watermarks

It is well known that adding physical watermarks can improve the detection rate by sacrificing control performances when the system is subject to replay attacks. However, existing watermarks are added based on time intervals. Since the watermarks will be added even when the system is under normal operation, resulting in excessive sacrifice of system performances.

In this section, we propose a stochastic event-based feedback physical watermark. To be more specific, at each step k, the computer center produces an i.i.d. variable ζ_k, which obeys a uniform distribution between [0,1] , and furthermore compares it with the function φ_k:

$$\varphi_k = \exp\{-\frac{1}{2} \sum_{i=k-\mathcal{K}+1}^{k} z_i^T Y z_i\}, \tag{21}$$

where $\mathcal{K}$ is the window size and $Y \succ 0$. The computer centre chooses to add a physical watermark if and only if $\varphi_k < \zeta_k$. Then γ_k is obtained by:

$$\gamma_k = \begin{cases} 0, & \text{with prob.} \varphi_k, \\ 1, & \text{with prob.} 1 - \varphi_k. \end{cases} \tag{22}$$

It should be pointed out that when the system is operating normal, the probability of adding a physical watermark is a constant. Furthermore, the probability of adding a watermark $1 - \varphi_k$ will increase when replay attacks exist, which will be proved in the following article.

Furthermore, we add a feedback channel to increase the watermark covariance (see Fig. 1). At present, the watermark can be expressed as $f(g_k)\gamma_k \Delta u_k$ where $f(\cdot) : \mathbb{R}^+ \to [0, \delta]$. The bound δ is to prevent an excessive watermark signal. Notably, if Y is large enough and $f(g_k) \equiv 1$, we will obtain the time-based watermarks.

In the rest of this article, both the $\mathscr{K}$ and $\mathcal{K}$ are set to 1. However, it is easy to extend our results to more general cases by state extension. We first focus on the situation where the system operates without replay attacks. The next theorem shows the probability of adding a watermark:

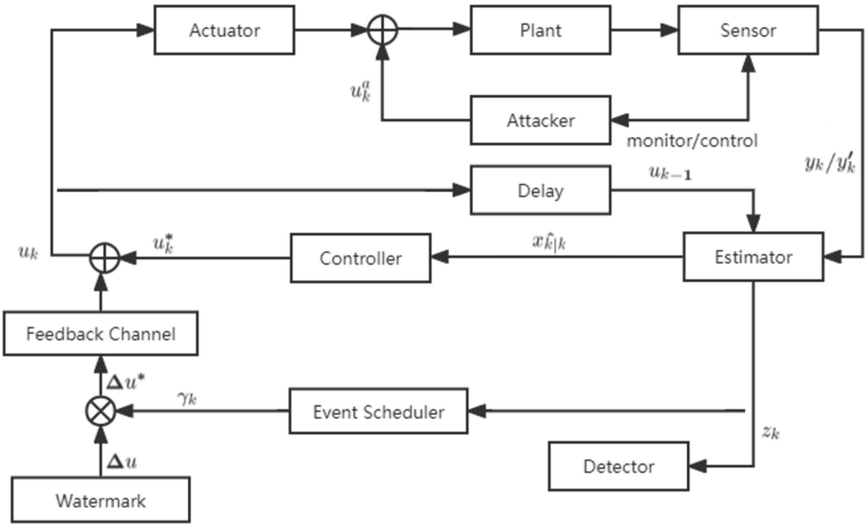

Fig. 1. System diagram

Theorem 1. *In the absence of replay attacks, the probability of adding a watermark is a constant*

$$\gamma = Pr(\gamma_k = 1) = 1 - \frac{1}{|I + Y\mathscr{P}|^{\frac{1}{2}}}. \tag{23}$$

Proof.

$$\begin{aligned} Pr(\gamma_k = 0) &= \mathbb{E}[\varphi_k] = \int_{\mathbb{R}^m} \varphi_k p(z_k) dz_k \\ &= \frac{1}{(2\pi)^{m/2}|\mathscr{P}|^{\frac{1}{2}}} \int_{\mathbb{R}^m} exp\{-\frac{1}{2} z_k^T Y z_k\} exp\{-\frac{1}{2} z_k^T \mathscr{P}^{-1} z_k\} dz_k \\ &= \frac{1}{|I + Y\mathscr{P}|^{\frac{1}{2}}}, \end{aligned} \tag{24}$$

Thus,

$$Pr(\gamma_k = 1) = 1 - \frac{1}{|I + Y\mathscr{P}|^{\frac{1}{2}}}. \tag{25}$$

From work [4], we know that

Lemma 3.

$$J^{'} = J + trace[(U + B^T S B)\mathscr{Q}], \tag{26}$$

where $J^{'}$ is the LQG performance after adding a physical watermark at each time step, and $\mathscr{Q}$ is the watermark covariance.

Then we can obtain the LQG performance with our watermarks by the following theorem:

Theorem 2. *The LQG performance with stochastic event-based feedback physical watermarks is:*

$$J^{'} \leq J + \Delta J, \tag{27}$$

where $\Delta J = \hat{q}trace[(U + B^T SB)\mathscr{Q}]$, *and* $\hat{q} = \mathbb{E}[f(g_0)^2]$.

Proof.

$$Cov(f(g_k)\gamma_k \Delta u_k) = \mathbb{E}[f(g_k)^2 \gamma_k]\mathscr{Q} \leq \mathbb{E}[f(g_k)^2]\mathscr{Q}. \tag{28}$$

Combining $Cov(f(g_k)\gamma_k \Delta u_k)$ with Lemma 3, we complete the proof.

We will verify the effectiveness of detecting replay attacks with our stochastic event-based feedback physical watermarks in the next section.

4 Detection Performance

We will consider the case where replay attacks exist in the following. According to the work [3], the attack model can be represented by the following system dynamics:

$$x'_{k+1} = Ax'_k + Bu'_k + w'_k, \tag{29}$$
$$y'_k = Cx'_k + v'_k, \tag{30}$$
$$\hat{x}'_{k+1|k} = A\hat{x}'_{k|k} + Bu'_k, \tag{31}$$
$$\hat{x}'_{k|k} = \hat{x}'_{k|k-1} + K(y'_k - \hat{x}'_{k|k-1}), \tag{32}$$
$$u'_k = L\hat{x}'_{k|k} + f(g'_k)\gamma'_k \Delta u'_k, \tag{33}$$

where γ'_k is the indication variable when the system operate without replay attacks and $\Delta u'_k \sim \mathcal{N}(0, \mathscr{Q})$. $\hat{x}_{k+1|k}$ and $\hat{x}'_{k+1|k}$ can be rewritten:

$$\hat{x}_{k+1|k} = \mathscr{A}\hat{x}_{k|k-1} + (A + BL)Ky'_k + Bf(g^a_k)\gamma_k \Delta u_k. \tag{34}$$
$$\hat{x}'_{k+1|k} = \mathscr{A}\hat{x}'_{k|k-1} + (A + BL)Ky'_k + Bf(g'_k)\gamma'_k \Delta u'_k. \tag{35}$$

Let $\hat{x}_{0|-1} - \hat{x}'_{0|-1} \triangleq \zeta$. Then the innovation of Kalman filter when replay attacks exist can be expressed as:

$$\begin{aligned} z^a_k &= y'_k - C\hat{x}_{k|k-1} \\ &= z'_k - C\mathscr{A}^k \zeta - C\sum_{i=0}^{k-1} \mathscr{A}^{k-i-1} B(f(g^a_i)\gamma_i \Delta u_i - f(g'_i)\gamma'_i \Delta u'_i), \end{aligned} \tag{36}$$

The next theorem shows the upper bound of $\mathbb{E}[f(g^a_k)^2]$ when the replay attacks exist.

Theorem 3. *Define* $q(\cdot) \triangleq f(\cdot)^2$. *We assume that* $q(\cdot)$ *is a monotonically increasing truncation function and the watermark is added at each time step* k *when the replay attacks exist, since using this method the* $\mathbb{E}[q(g_k^a)]$ *is the largest, and the upper bound of* $\mathbb{E}[q(g_k^a)]$ *is* $\bar{q}$ *and* $\bar{q} \geq \hat{q}$.

Proof. It is worth noted that Δu_i is independent of $f(g_j^a)$ when $j \leq i$ and the vectors of the virtual system for all k. The proof can refer to the proof of Theorem 3 in work [16], so it is omitted here.

The following theorem gives the upper and lower bounds of the probability of adding a physical watermark when the replay attacks exist.

Theorem 4. *The probability of adding a watermark will increase when replay attacks exist. The bounds of* $Pr(\gamma_k = 1)$ *are given as follows:*

$$\lim_{k\to\infty} Pr(\gamma_k = 1) \leq 1 - \frac{1}{|I + Y\mathscr{H}|^{\frac{1}{2}}}, \tag{37}$$

$$\lim_{k\to\infty} Pr(\gamma_k = 1) \geq \gamma, \tag{38}$$

where

$$\mathscr{H} = \mathscr{P} + C\mathscr{M}C^T + C\mathscr{N}C^T, \tag{39}$$

$\mathscr{M}$ *and* $\mathscr{N}$ *are the solutions of the following Lyapunov equations, respectively*

$$\mathscr{M} - \bar{q}B\mathscr{Q}B^T = \mathscr{A}\mathscr{M}\mathscr{A}^T, \tag{40}$$

$$\mathscr{N} - \hat{q}B\mathscr{Q}B^T = \mathscr{A}\mathscr{N}\mathscr{A}^T. \tag{41}$$

Proof. Define $\tilde{\gamma_k} \triangleq (\gamma_1, \gamma_1', ..., \gamma_k, \gamma_k')$. It can be seen from (36) that when k is large enough, the expectation of z_k^a will converge to 0, so $z_k^a \sim \mathcal{N}(0, \Pi_k(\tilde{\gamma}_{k-1}))$, where

$$\begin{aligned}\Pi_k(\tilde{\gamma}_{k-1}) = \mathscr{P} + C\sum_{i=0}^{k-1}\{&\mathbb{E}[q(g_i^a)\gamma_i]\mathscr{A}^{k-i-1}B\mathscr{Q}B^T(\mathscr{A}^{k-i-1})^T \\ &+ \mathbb{E}[q(g_i')\gamma_i']\mathscr{A}^{k-i-1}B\mathscr{Q}B^T(\mathscr{A}^{k-i-1})^T\}C^T.\end{aligned} \tag{42}$$

The probability of not adding a watermark:

$$\begin{aligned}Pr(\gamma_k = 0) &= \mathbb{E}_{\tilde{\gamma}_{k-1}}\left[\int_{\mathbb{R}^m} \varphi_k p(z_k^a) dz_k^a\right] \\ &= \mathbb{E}_{\tilde{\gamma}_{k-1}}\left[\frac{1}{(2\pi)^{m/2}|\Pi_k(\tilde{\gamma}_{k-1})|^{\frac{1}{2}}}\right. \\ &\left.\times \int_{\mathbb{R}^m} exp\{-\frac{1}{2}z_k^{aT}Yz_k^a\}exp\{-\frac{1}{2}z_k^{aT}\Pi_k(\tilde{\gamma}_{k-1})^{-1}z_k^a\}dz_k^a\right] \\ &= \mathbb{E}_{\tilde{\gamma}_{k-1}}\left[\frac{1}{|I + Y\Pi_k(\tilde{\gamma}_{k-1})|^{\frac{1}{2}}}\right].\end{aligned} \tag{43}$$

Combining (42) and Theorem 3, we have

$$\begin{aligned} &\Pi_k(\tilde{\gamma}_{k-1}) \geq \mathscr{P} \\ &\Pi_k(\tilde{\gamma}_{k-1}) \leq \mathscr{P} + C\sum_{i=0}^{k-1}\{\bar{q}\mathscr{A}^{k-i-1}B\mathscr{Q}B^T(\mathscr{A}^{k-i-1})^T \\ &\qquad + \hat{q}\mathscr{A}^{k-i-1}B\mathscr{Q}B^T(\mathscr{A}^{k-i-1})^T\}C^T = \mathscr{H}. \end{aligned} \tag{44}$$

Then we can acquire that

$$\lim_{k\to\infty} Pr(\gamma_k = 0) \leq \frac{1}{|I + Y\mathscr{P}|^{\frac{1}{2}}} \tag{45}$$

$$\lim_{k\to\infty} Pr(\gamma_k = 0) \geq \frac{1}{|I + Y\mathscr{H}|^{\frac{1}{2}}}. \tag{46}$$

which finishes the proof.

In the next theorem, we analyze the performance of the χ^2 detector based on stochastic event-based feedback physical watermarks. We will show the utility of our watermarks by proving the improvement in detection rate.

Theorem 5. *When the system operates without replay attacks,*

$$\mathbb{E}[g_k] = m. \tag{47}$$

When the replay attacks exist,

$$\begin{aligned} \lim_{k\to\infty} \mathbb{E}[g_k] &\leq m + trace(C^T\mathscr{P}^{-1}C\mathscr{M}) \\ &\quad + trace(C^T\mathscr{P}^{-1}C\mathscr{N}), \end{aligned} \tag{48}$$

$$\lim_{k\to\infty} \mathbb{E}[g_k] \geq m. \tag{49}$$

Proof. When the system operates without replay attacks, it is trivial to prove

$$\mathbb{E}[z_k^T\mathscr{P}^{-1}z_k] = trace(\mathscr{P}^{-1}\mathbb{E}(z_k z_k^T)) = m. \tag{50}$$

When the replay attacks exist, from (44)

$$\mathscr{P} \leq \lim_{k\to\infty} Cov(z_k^a) \leq \mathscr{P} + C\mathscr{M}C^T + C\mathscr{N}C^T. \tag{51}$$

Hence, we can obtain that

$$\begin{aligned} &\lim_{k\to\infty} \mathbb{E}[z_k^{aT}\mathscr{P}^{-1}z_k^a] = \lim_{k\to\infty} trace(\mathscr{P}^{-1}\mathbb{E}(z_k^a z_k^{aT})) \\ &\leq m + trace(C^T\mathscr{P}^{-1}C\mathscr{M}) + trace(C^T\mathscr{P}^{-1}C\mathscr{N}), \end{aligned} \tag{52}$$

$$\begin{aligned} &\lim_{k\to\infty} \mathbb{E}[z_k^{aT}\mathscr{P}^{-1}z_k^a] = \lim_{k\to\infty} trace(\mathscr{P}^{-1}\mathbb{E}(z_k^a z_k^{aT})) \\ &\geq m. \end{aligned} \tag{53}$$

5 Numerical Example

In this section, we will demonstrate the effectiveness of stochastic event-based feedback physical watermarks by comparing with the ordinary watermarks [3]. For simplicity, we analyze the detection of replay attacks on the following system: $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $Q = 0.8I$, $R = 1$, $W = I$, $U = 1$.

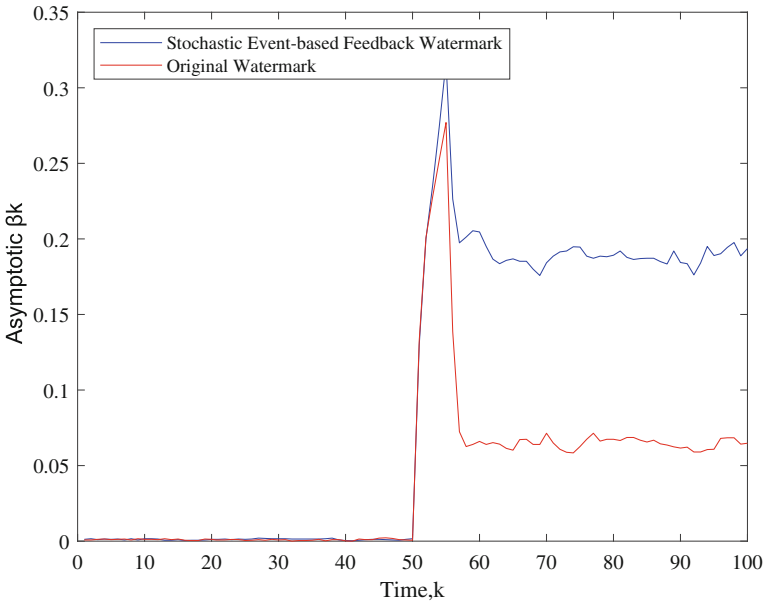

Fig. 2. Detection rates under replay attacks

The eigenvalues of $\mathscr{A}$ can be calculated as -0.339 and -0.105. Hence the system is vulnerable to replay attacks. The window sizes $\mathscr{K}$ (the χ^2 detector) and $\mathcal{K}$ (the event trigger) are set to be 5 and $f(g_k) = \sqrt{g_k}$ with upper bound $\delta = 5$, which means that the output of the χ^2 detector can be used directly. In this case, $\hat{h} = m\mathscr{T} = 5$.

Based on the above parameters, it is easy to get that $J = 23.1$. In order to better compare with the original watermarks, we set the LQG cost J' the same as $J' = 1.47J$, and the detection rates of χ^2 detector for different watermarks will be displayed. Attackers record the sensor readings from time 1 to time 50 and replay them from time 51 to time 100. In addition, the false alarm rate $\alpha_k = 0.001$. Each result takes an average of 5000 experiments. Firstly, we set $Y = 1.25\mathscr{P}^{-1}$. The covariance of our watermarks $\mathscr{Q} = 0.41$ and the covariance of ordinary watermarks $\mathscr{Q} = 1.922$. It is worth pointing out that we use this form of Y since the event scheduler can directly use g_k. We can know that when using

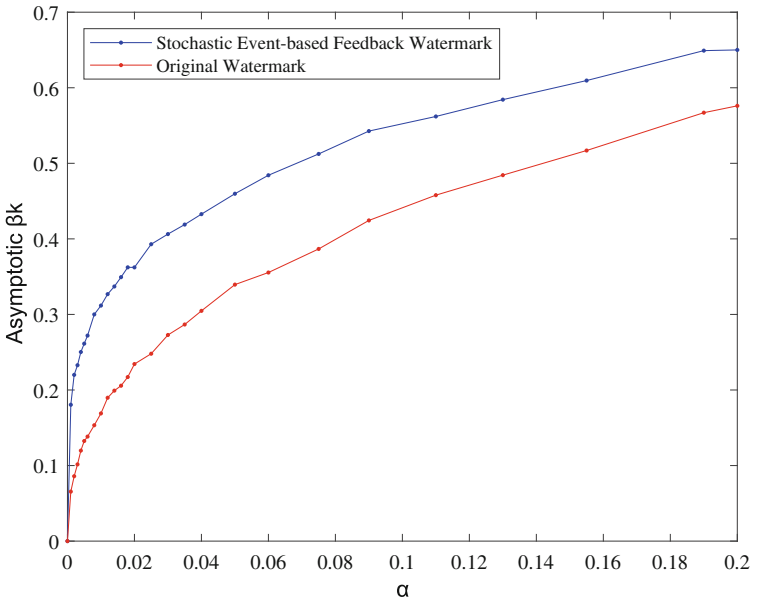

Fig. 3. Receiver operating characteristic (ROC) curves

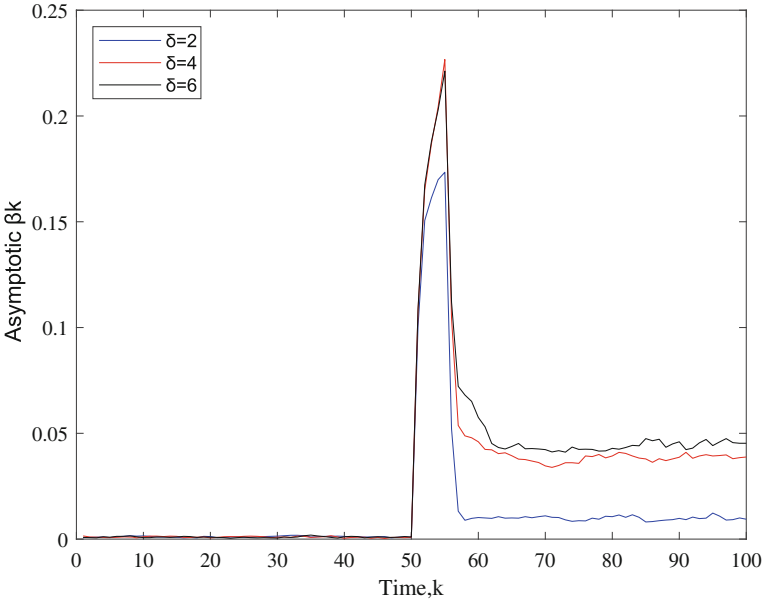

Fig. 4. Relationship between bound δ and asymptotic β_k

the stochastic event-based feedback physical watermarks, the detection rate is higher under the same performance loss according to Fig. 2. Then the ROC curves for different watermark approaches are drawn in Fig. 3. From Fig. 2 and Fig. 3, it can be immediately concluded that our watermarks are more effective than the ordinary watermarks.

Finally, we intend to study the impact of the bound δ. We set $Y = 1.25\mathscr{P}^{-1}$ and $\mathscr{Q} = 0.2$ with $\delta = 2, 4, 6$, respectively. Figure 4 shows that a larger δ can improve the detection rate. However, it is also easy to undermine the certainty of the system.

6 Conclusion

A novel stochastic event-based feedback watermark is proposed in this paper to defend against replay attacks. Firstly, the probability of adding a watermark and the LQG performance loss when the system operates without replay attacks are analyzed. Secondly, the asymptotic upper and lower bounds of the probability of adding a stochastic event-based feedback watermark in the presence of replay attacks are discussed, and it is proved that the probability of adding a physical watermark will increase when replay attacks exist. Furthermore, the utility of our watermarks is proved by calculating the detection rate of the χ^2 detector. Finally, several numerical simulations illustrate that our method is more effective than the ordinary watermarks. In future work, we will consider parameter optimization and feedback function design.

References

1. Zhang, X.-M., Han, Q.-L., Yu, X.: Survey on recent advances in networked control systems. IEEE Trans. Industr. Inf. **12**(5), 1740–1752 (2015)
2. Cardenas, A.-A., Amin, S., Sastry, S.: Secure control: Towards survivable cyber-physical systems. In: 28th International Conference on Distributed Computing Systems Workshops, pp. 495–500 (2008)
3. Mo, Y., Sinopoli, B.: Secure control against replay attacks. In: 47th annual Allerton Conference On Communication, Control, and Computing (Allerton), pp. 911–918 (2009)
4. Mo, Y., Chabukswar, R., Sinopoli, B.: Detecting integrity attacks on SCADA systems. IEEE Trans. Control Syst. Technol. **22**(4), 1396–1407 (2013)
5. Mo, Y., Weerakkody, S., Sinopoli, B.: Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. IEEE Control Syst. Mag. **35**(1), 93–109 (2015)
6. Weerakkody, S., Ozel, O., Sinopoli, B.: A Bernoulli-Gaussian physical watermark for detecting integrity attacks in control systems. In: 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 966–973 (2017)
7. Liu, H., Yan, J., Mo, Y.: An on-line design of physical watermarks. In: IEEE Conference on Decision and Control (CDC), pp. 440–445 (2018)
8. Fang, C., Qi, Y., Cheng, P.: Cost-effective watermark based detector for replay attacks on cyber-physical systems. In: 11th Asian Control Conference (ASCC), pp. 940–945 (2017)

9. Ferrari, R.-M., Teixeira, A.-M.: Detection and isolation of replay attacks through sensor watermarking. IFAC-PapersOnLine **50**(1), 7363–7368 (2017)
10. Miao, F., Pajic, M., Pappas, G.-J.: Stochastic game approach for replay attack detection. In: 52nd IEEE Conference on Decision and Control, pp. 1854–1859 (2013)
11. Miao, F., Zhu, Q., Pajic, M.: A hybrid stochastic game for secure control of cyber-physical systems. Automatica **93**, 55–63 (2018)
12. Shi, L., Johansson, K.-H., Qiu, L.: Time and event-based sensor scheduling for networks with limited communication resources. IFAC Proceed. Vol. **44**(1), 13263–13268 (2011)
13. Han, D., Mo, Y., Wu, J.: Stochastic event-triggered sensor schedule for remote state estimation. IEEE Trans. Autom. Control **60**(10), 2661–2675 (2015)
14. Bertsekas, D.: Dynamic programming and optimal control: Volume I. Athena scientific (2012)
15. Mehra, R.-K., Peschon, J.: An innovations approach to fault detection and diagnosis in dynamic systems. Automatica **7**(5), 637–640 (1971)
16. Zhao, X., Liu, L., Karimi, H.-R.: Detection against replay attack: a feedback watermark approach. In: International Summit Smart City 360°, pp. 702–715 (2022)