European Journal of Control 62 (2021) 57–62

Contents lists available at ScienceDirect

European Journal of Control

journal homepage: www.elsevier.com/locate/ejcon

Physical watermarking for replay attack detection in continuous-time systems

Bahram Yaghooti[a,*], Raffaele Romagnoli[b], Bruno Sinopoli[a]

[a] Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA
[b] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 20 April 2021
Revised 10 June 2021
Accepted 25 June 2021
Available online 14 July 2021

Recommended by Prof. T Parisini

Keywords:
Cyber-physical systems
Physical watermarking
Replay attack
Security

ABSTRACT

Physical watermarking is a well established technique for replay attack detection in cyber-physical systems (CPSs). Most of the watermarking methods proposed in the literature are designed for discrete-time systems. In general real physical systems evolve in continuous time. In this paper, we analyze the effect of watermarking on sampled-data continuous-time systems controlled via a Zero-Order Hold. We investigate the effect of sampling on detection performance and we provide a procedure to find a suitable sampling period that ensures detectability and acceptable control performance. Simulations on a quadrotor system are used to illustrate the effectiveness of the theoretical results.

1. Introduction

Cyber-Physical Systems (CPSs) integrate communication, computation, and control into physical world. CPSs play a crucial role in the design of efficient and sustainable services that are pillars of modern societies, such as energy delivery, transportation, health care, and water distribution [15]. Their safety and security represent one of the main design challenges, as their heterogeneous and distributed nature makes CPSs vulnerable to a multitude of cyber-attacks [16,25].

The problem of cyber-attacks in networked control systems has been studied comprehensively in previous research [6,11–13,22]. In this paper our focus is on the analysis and detection of *replay attacks* in control systems [4,14]. In such an attack, a cybercriminal eavesdrops on a network, and then fraudulently delays or resends old observations to misdirect the receiver into thinking that the system is behaving normally while carrying out their attack. One of the characteristics of replay attack which makes it simple to implement is that it can be used by a hacker without advanced knowledge of the system or skills to decrypt messages. This type of attack can be successful just by repeating a set of recorded data.

The first model of replay attacks on control systems together with a proposed countermeasure was introduced by Mo and Sinopoli [18] and refined in subsequent papers [19,20]. The basic idea revolves around the use of physical watermarking, a secret noisy control input added to an intended control input and aimed at authenticating the received observation. This framework was developed for discrete-time systems, and its performance has been studied extensively in the literature [8,10,23].

In general, many physical processes are continuous-time and controlled by sampling outputs and using a Zero-Order-Hold (Z.O.H) method for control. In this paper, we investigate the application of the watermarking framework to continuous-time systems and analyze the effect of sampling period on its performance. Specifically, while it is known that decreasing sampling period improves the performance of the controller, we will show that sampling also affects the performance of the detector. In this paper we explore this tradeoff to design an optimal sampling period. Finally, to illustrate the theoretical results, we apply the proposed methodology to a quadrotor hovering around an equilibrium point.

The rest of the paper is organized as follows: In Section 2, the discretization of a linear continuous-time stochastic systems is reviewed. Section 3 provides basic concepts of the watermarking framework. Section 4 investigates the effect of the sampling period on the controller and detector. It is shown that in the case of physical watermarking, the sampling period T becomes a design parameter alongside the covariance of the watermarking signal. We therefore generalize the design of the watermarking signal in [20] by defining a new optimization problem where we jointly design both the covariance of the watermarking and the sampling period, with appropriate constraints on the loss of performance and the maximum allowable sampling period as per [5]. To eval-

* Corresponding author.
E-mail address: byaghooti@wustl.edu (B. Yaghooti).

https://doi.org/10.1016/j.ejcon.2021.06.012
0947-3580/

uate the theoretical results, the watermarking method is applied to a quadrotor, and simulation results are provided in Section 5. Finally, a brief conclusion is presented in the last section.

2. System description

Consider the following linear continuous-time stochastic system:

$$\dot{x}(t) = Ax(t) + Bu(t) + w(t) \tag{1}$$

$$y(t) = Cx(t) + v(t), \tag{2}$$

where $x \in \mathbb{R}^n$ is the system state vector; $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$, and $C \in \mathbb{R}^{m\times n}$ are known and constant matrices; $u(t) \in \mathbb{R}^p$ and $y(t) \in \mathbb{R}^m$ are the input and output of the system, respectively; and $w(t) \in \mathbb{R}^n$ and $v(t) \in \mathbb{R}^m$ are the process and measurement noises, respectively. It is assumed that $w(t)$ and $v(t)$ are zero-mean Gaussian white noises,

$$\mathbb{E}[w(t)] = 0 \tag{3}$$

$$\mathbb{E}[v(t)] = 0 \tag{4}$$

$$\mathbb{E}[w(t)w^\top(s)] = Q\delta(t-s) \tag{5}$$

$$\mathbb{E}[v(t)v^\top(s)] = R\delta(t-s), \tag{6}$$

where $\delta(\cdot)$ is Dirac delta function, and $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are known matrices [2].

It is assumed that the process and measurement noises are independent of the previous and current state and independent of each other:

$$\mathbb{E}[w(t)x^\top(s)] = 0 \quad \text{for } t \geq s \tag{7}$$

$$\mathbb{E}[v(t)x^\top(s)] = 0 \quad \text{for } t \geq s \tag{8}$$

$$\mathbb{E}[w(t)v^\top(s)] = 0 \quad \text{for all } t, s \geq 0. \tag{9}$$

Physical watermarking has been used for replay attack detection in control systems. As this framework is developed for discrete-time systems, we first need to discretize system (1). To this end, the control input will be applied using Zero-Order-Hold (ZOH) method with sampling period T, so the controller will produce a piece-wise constant command between sampling periods,

$$u(t) = u_k, \quad kT \leq t < (k+1)T, \quad k = 0, 1, \ldots. \tag{10}$$

By integrating both sides of (1), the corresponding discrete-time system can be obtained as follows:

$$x_{k+1} = A_d x_k + B_d u_k + w_k, \tag{11}$$

where A_d, B_d, and w_k are defined as

$$A_d = e^{AT} = \sum_{m=0}^{\infty} \frac{A^m T^m}{m!} \tag{12}$$

$$B_d = \left(\int_0^T e^{A(T-\tau)}d\tau\right)B = \sum_{m=0}^{\infty} \frac{A^m B T^{m+1}}{(m+1)!} \tag{13}$$

$$w_k = \int_{kT}^{(k+1)T} e^{A((k+1)T-\tau)} w(\tau)d\tau. \tag{14}$$

The discretized Gaussian noise w_k has zero mean and convariance

$$Q_d = \int_0^T e^{A(T-\tau)} Q e^{A^\top(T-\tau)} d\tau. \tag{15}$$

The output of the discrete-time system can be obtained from Eq. (2) in the following form

$$y_k = Cx_k + v_k, \tag{16}$$

where v_k is zero-mean Gaussian measurement noise with the covariance R_d, i.e.

$$v_k \sim \mathcal{N}(0, R_d). \tag{17}$$

Differently from (15), the covariance matrix R_d has to be approximated as

$$R_d \approx \frac{R}{T}, \tag{18}$$

where R_d is the covariance matrix of the sampled noise signal v_k. Approximation (18) makes use of the rectangular function of amplitude $1/T$ to approximate the Dirac delta function $\delta(\cdot)$ in (6). In this way, the area under the rectangular function is R, and for the sampling period T that goes to zero, R_d converges to R. More details about the sampling of continuous-time stochastic systems are as shown in [1,7].

3. Review of watermarking framework in discrete-time systems

In this section, a brief description is presented for the watermarking framework which has been developed for replay attack detection in discrete-time control systems [19]. We assume that a hacker wants to corrupt the system defined in (11) and equipped with a Kalman filter and an LQG controller. We consider two main assumptions: (1) The attacker has access to all the sensors data; (2) The attacker can inject any control input to the system. To detect this attack, a watermarking signal is added to the normal control input. By doing so, we can distinguish between a normally operating system, which is driven by the current watermarks, and an attacked system, which is driven by a previous sequence of watermarks.

3.1. Kalman filter

It is well known that the Kalman filter provides the optimal state estimate $\hat{x}_{k|k}$ for system (11), as it provides the minimum variance unbiased estimate of the state x_k.

$$\begin{aligned}
&\hat{x}_{0|-1} = \bar{x}_0, \quad P_{0|-1} = \Sigma \\
&\hat{x}_{k+1|k} = A_d\hat{x}_k + B_d u_k \\
&P_{k+1|k} = A_d P_k A_d^\top + Q_d \\
&K_k = P_{k|k-1}C^\top(CP_{k|k-1}C^\top + R_d)^{-1} \\
&\hat{x}_k = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \\
&P_k = P_{k|k-1} - K_k C P_{k|k-1}
\end{aligned} \tag{19}$$

When the usual conditions provided by Kalman, the estimator gain converges to its steady-state value, and in most of applications, this convergence occurs in a few steps. Therefore, we can write state error covariance, P, and Kalman gain, K, as follows:

$$P \triangleq \lim_{k\to\infty} P_{k|k-1}, \quad K \triangleq PC^\top(CPC^\top + R_d)^{-1}. \tag{20}$$

We assume the system to be in steady state. By considering $\Sigma = P$ as the initial condition, one can rewrite the Kalman filter as a fixed gain estimator in the following form:

$$\begin{aligned}
&\hat{x}_{0|-1} = \bar{x}_0, \quad \hat{x}_{k+1|k} = A_d\hat{x}_k + B_d u_k \\
&\hat{x}_k = \hat{x}_{k|k-1} + K(y_k - C\hat{x}_{k|k-1}).
\end{aligned} \tag{21}$$

3.2. Linear quadratic Gaussian (LQG) control

In this section, an LQG control scheme is designed such that minimize the following infinite-horizon objective function

$$J = \lim_{N\to\infty} \mathbb{E}\frac{1}{N}\Big[\sum_{k=0}^{N-1}(x_k^\top W x_k + u_k^\top U u_k)\Big], \tag{22}$$

where W and U are positive semi-definite matrices. It is known that the above optimization problem yields the following fixed gain control input

$$u_k = u_k^* = -(B_d^\top S B_d + U)^{-1} B_d^\top S A_d \hat{x}_{k|k}, \tag{23}$$

where S can be obtained by solving the well-known infinite-horizon Riccati equation

$$S = A_d^\top S A_d + W - A_d^\top S B_d (B_d^\top S B_d + U)^{-1} B_d^\top S A_d. \tag{24}$$

By defining $L = -(B_d^\top S B_d + U)^{-1} B_d^\top S A_d$, the LQG controller can be rewritten as $u_k^* = L\hat{x}_{k|k}$.

3.3. χ^2TEXT failure detector

The χ^2 detector has been widely used to detect anomalies in control systems. Principle idea of this detector is based on the probability distribution of the residual of Kalman filter.

Theorem 3.1. *[17] For the discrete-time system defined by (11) with Kalman filter and LQG controller, the residuals $y_i - C\hat{x}_{i|i-1}$ of Kalman filter are i.i.d. Gaussian distributed with zero mean and covariance $\mathscr{P}$, where $\mathscr{P} = CPC^\top + R_d$.*

By using Theorem 3.1, when the system is in normal condition, the probability to get the sequence $y_{k-\mathscr{T}+1}, \dots, y_k$ can be obtained as follows

$$P(y_{k-\mathscr{T}+1}, \dots, y_k) = \left[\frac{1}{(2\pi)^{N/2}|\mathscr{P}|}\right]^{\mathscr{T}} \exp\Big(-\frac{1}{2}g_k\Big), \tag{25}$$

where $\mathscr{T}$ is the window size of the detection, and g_k is defined as

$$g_k = \sum_{i=k-\mathscr{T}+1}^{k} (y_i - C\hat{x}_{i|i-1})^\top \mathscr{P}^{-1} (y_i - C\hat{x}_{i|i-1}). \tag{26}$$

When the probability of detection is low, one conclude that there is an anomaly in the system, but in order to apply χ^2 detector, we do not need to calculate this probability. when the system is in normal condition, g_k has a χ^2 distribution with $m\mathscr{T}$ degrees of freedom. Then, One can use Eq. (26) to detect any failure in the system. Therefore, the χ^2 detector can be rewritten as

$$g_k \lessgtr threshold, \tag{27}$$

where threshold is chosen for a specific false alarm probability. If g_k is greater than the threshold, then the detector will trigger an alarm.

3.4. Detection of replay attack in control systems

In the replay attack, the attacker resends a set of data which is recorded for a period of time. Specifically, the attacker records the measurements from time k' to $k'+N$, and replaces the actual measurements from $k \geq k'+N+1$. We indicate the virtual output as $y'_k \triangleq y_{k-N+1}$ for $k \geq k'+N+1$. Since,

$$y'_k = Cx_{k-N+1} + v_{k-N+1}.$$

we also define $x'_k \triangleq x_{k-N+1}$, $v'_k \triangleq v_{k-N+1}$ and $u'_k \triangleq u_{k-N+1}$. At this point we can define the following virtual system that describes the "shifted" dynamics of the system and observer.

$$\begin{aligned} x'_{k+1} &= A_d x'_k + B_d u'_k, \quad y'_k = Cx'_k + v'_k \\ \hat{x}'_{k+1|k} &= A_d \hat{x}'_{k|k} + B_d u'_k \\ \hat{x}'_{k+1|k+1} &= \hat{x}'_{k+1|k} + K(y'_k - \hat{x}'_{k+1|k}) \\ u'_k &= L\hat{x}'_{k|K}. \end{aligned} \tag{28}$$

During the replay attack, the residuals (26) are obtained replacing y_k with y'_k, and the one-step prediction of the state $\hat{x}_{k|k-1}$ assumes the following form

$$\hat{x}_{k+1|k} = (A_d + B_d L)(I - KC)\hat{x}_{k|k-1} + (A_d + B_d L)Ky'_k. \tag{29}$$

By using the one-step prediction of the state of the virtual system (28), defined as

$$\hat{x}'_{k+1|k} = (A_d + B_d L)(I - KC)\hat{x}'_{k|k-1} + (A_d + B_d L)Ky'_k, \tag{30}$$

the residuals (26) can be rewritten as

$$\begin{aligned} g_k = \sum_{i=k-\mathscr{T}+1}^{k} \Big[& (y'_i - C\hat{x}'_{i|i-1})^\top \mathscr{P}^{-1} (y'_i - C\hat{x}'_{i|i-1}) \\ & + 2(y'_i - C\hat{x}'_{i|i-1})^\top \mathscr{P}^{-1} C\mathscr{A}^i \zeta + \zeta^\top (\mathscr{A}^i)^\top C^\top \mathscr{P}^{-1} C\mathscr{A}^i \zeta \Big], \end{aligned} \tag{31}$$

where $\mathscr{A} \triangleq (A_d + B_d L)(I - KC)$ and $\zeta \triangleq \hat{x}_{0|-1} - \hat{x}'_{0|-1}$.

To evaluate the performance of designed χ^2 detector, we need to consider two cases.

1. $\mathscr{A}$ is stable: In this case, the second and third terms in (31) will converge to zero. Hence, g_k for the main and virtual system has the same distribution. Then, the detector is completely useless to detect any replay attack in the control system.
2. $\mathscr{A}$ is unstable: Any replay attack which is applied for a long time can be detected by the χ^2 detector, because g_k will soon become unbounded, and by comparing g_k for the main and virtual system, the replay attack will be detected.

We conclude that the χ^2 detector is useful only for an unstable $\mathscr{A}$. To be able to detect replay attack when $\mathscr{A}$ is stable, the control input is redesigned by adding an authentication signal. Let us define the new controller in the following form:

$$u_k = u_k^* + \Delta u_k, \tag{32}$$

where u_k^* is the LQG optimal control which is defined by (23), and Δu_k is an authentication signal added to the optimal control to be able to detect replay attack. The sequence Δu_k is drawn from an i.i.d. Gaussian distribution with zero mean and covariance $\mathscr{Q}$, and independent of u_k^*. By adding this authentication signal, the control input will not be the optimal one. However, it will help us to detect replay attack. In other words, the watermarking framework sacrifice the control performance to be able to detect the attack.

Theorem 3.2. *[20] After adding the authentication signal to the optimal control, the LQG performance is given by*

$$J' = J + trace[(U + B_d^\top S B_d)\mathscr{Q}]. \tag{33}$$

The following theorem represents performance of the χ^2 detector in presence of replay attack.

Corollary 3.3. *[20] In the absence of an attack, the expectation of g_k in the χ^2 detector is*

$$\mathbb{E}[g_k] = m\mathscr{T}. \tag{34}$$

Under attack, the asymptotic expectation becomes

$$\lim_{k\to\infty} \mathbb{E}[g_k] = m\mathscr{T} + 2trace(C^\top \mathscr{P}^{-1} C\mathscr{U})\mathscr{T}, \tag{35}$$

where $\mathscr{U}$ is the solution to the following equation

$$\mathscr{U} = \sum_{i=0}^{\infty} \mathscr{A}^i B_d \mathscr{Q} B_d^\top (\mathscr{A}^i)^\top. \tag{36}$$

The difference in the expectations of g_k with and without attack proves that the detection rate does not converge to the false alarm rate.

Remark 3.4. The physical watermarking framework has been developed for systems controlled by an LQG controller. The focus of this paper is not devoted to study the effects of different kind of controllers on this framework, but due to the connections between LQG, the generalized $\mathcal{H}_2$ control, and $\mathcal{H}_\infty$ control [21,24], our solution can be extended to these other kind of controllers.

4. Watermarking in continuous-time systems

In this section, the effect of the sampling period on performance of the watermarking framework is studied. In digital signal processing, the sampling theorem states that to reconstruct an unknown band-limited signal from discretized version of that signal, the sampling rate must be at least twice as high as the highest frequency in the signal. In digital control, this theorem is applied to a feedback controller. Thus, based on this theorem the sampling rate must be at least twice the required closed-loop bandwidth of the system. In most of applications, to get an appropriate time response for the control system, this sampling rate would be inadequate. Moreover, one of the most important concepts which should be considered is the delay between a command input and the system response to the command input. This delay should be reduced as much as possible. In order to confront these issues, Franklin et al. [5] proposed that the sampling rate should be at least 20 times the required closed-loop bandwidth of the system. This is a lower bound for sampling rate. Therefore, this shows that decreasing sampling period will increase the control performance. However, to evaluate the performance of watermarking framework, we need to analyze performance of controller and detector simultaneously. In the next subsection, we will analyze the effect of sampling period on the χ^2 detector performance.

4.1. Effect of sampling period on the χ^2 TEXT detector

Sampling period affects not only the control system response but also performance of the detector. As it is mentioned, if the sampling period decreases, time response of the control system will improve. The effect of small sampling period on performance of the χ^2 detector is studied in the following lemma and corollary.

Lemma 4.1. *Consider system (1) with the LQG controller (23) designed on the sampled data system (11), assuming that $\mathscr{U}$ be bounded for any $T > 0$, then as T goes to zero, the detector (27) is not able to detect any replay attack.*

Proof. By using a very small sampling period, i.e. $T \to 0$, the covariance of the measurement noise in the discretized system can be approximated by $R_d \approx R/T$. By substituting this approximation in the covariance of the residual of Kalman filter, its covariance can be rewritten as

$$\mathscr{P} = CPC^\top + \frac{R}{T}. \tag{37}$$

All the terms in the Taylor expansions of A_d, B_d, and Q_d are constant or $\mathcal{O}(T)$.[1] Because of the presence of $(CP_{k|k-1}C^\top + R/T)^{-1}$ in K_k and by considering the fact that we use very small sampling period, this term can be considered an $\mathcal{O}(T)$. Given that the Kalman filter does not have any term $1/T$, the matrix P does not contain $1/T$. Since we assumed that the sampling period is very small, the second term in (37) will be the dominant term. Thus, the covariance of the residuals of the Kalman filter and its inverse can be approximated as below

$$\mathscr{P} \approx \frac{R}{T}, \quad \mathscr{P}^{-1} \approx TR^{-1}. \tag{38}$$

By substituting $\mathscr{P}^{-1}$ from (38) into (35), the difference in the expectation of g_k with and without attack can be written as

$$\mathbb{E}[\Delta g_k] = 2trace(C^\top R^{-1} C\mathscr{U})\mathscr{T}T. \tag{39}$$

Therefore, as the sampling period tends to zero, the difference in the expectation of g_k with and without attack tends to zero, and according to (35) of Corollary 1, the detector is not able to detect any replay attack. □

Remark 4.2. As the sampling period goes to zero, the replay attack detection rate decreases. On the other hand, by increasing the sampling period, $\mathbb{E}[\Delta g_k]$ decreases due to the degradation of the control performance which results in larger steady state error and consequently in a lower detection rate. Then, there is an optimal value of the sampling period that maximizes detection performance.

In the next subsection, we proposed an optimization to manage the trade-off between detection rate and control performance.

4.2. Optimization

We have shown in the previous subsection that the sampling period is a compromise between the performance of the controller and detector. Therefore, to find the optimal value of the sampling period and covariance of watermarking signal, we have to solve the following optimization problem.

$$\begin{aligned} \max_{\mathscr{Q},T} \quad & 2trace(C^\top \mathscr{P}^{-1} C\mathscr{U})\mathscr{T} \\ \text{subject to} \quad & trace[(U + B_d^\top S B_d)\mathscr{Q}] < \mu \\ & 0 < T \le \bar{T} \\ & \mathscr{U} - B_d \mathscr{Q} B_d^\top = \mathscr{A}\mathscr{U}\mathscr{A}^\top, \end{aligned} \tag{40}$$

where $trace[(U + B_d^\top S B_d)\mathscr{Q}]$ is the extra cost introduced by the authentication signal according to Theorem 3.2, and μ is the corresponding upper bound. The sampling period should be chosen small enough such that the approximation $R_d = \frac{R}{T}$ be still valid. In general, there is not a rigorous method to define the upper bound for this approximation, and it should be calculated for any specific problem. The best way to find the upper bound is experimental results. The sampling period T is also bounded by $\bar{T}$ which has to be designed to prevent aliasing and guarantee specific control performance in terms of system response. To avoid the aliasing, from the sampling theorem, the sampling frequency must be at least twice the required cutoff frequency of the system. This provides the fundamental lower bound on the sampling frequency. However, in most of the applications, this frequency would be too slow for an acceptable time response. A common practice in this situation is to set the sampling frequency at least twenty times more than the bandwidth of the system [5]. The minimum value of these two criteria is the upper bound of the sampling period, $\bar{T}$.

Since all the parameters of the system, controller, and detector are functions of the sampling period the optimization problem (40) may be hard to solve. To simplify this procedure we fix the sampling period T in order to compute the following optimization problem

$$\begin{aligned} \max_{\mathscr{Q}} \quad & 2trace(C^\top \mathscr{P}^{-1} C\mathscr{U})\mathscr{T} \\ \text{subject to} \quad & trace[(U + B_d^\top S B_d)\mathscr{Q}] < \mu \\ & \mathscr{U} - B_d \mathscr{Q} B_d^\top = \mathscr{A}\mathscr{U}\mathscr{A}^\top. \end{aligned} \tag{41}$$

[1] $g(T) = \mathcal{O}(T)$ as $T \to 0$: "asymptotically g goes to zero at least as fast as T", or more formally: $\exists K \ge 0$ s.t. $\left|\frac{g(T)}{T}\right| \le K$ as $T \to 0$.

Then, we iterate the computation of (41) for different values of T. Bisection methods can be used to find the optimal sampling period T. To have a fair comparison, a constant value μ is used as the extra cost for all the sampling period values.

Remark 4.3. The optimization problem (41) can be rewritten in another way. We can maximize the increase (Δg_k) in the expected value of the quadratic residues in case of an attack, while constraining the LQG performance loss (ΔJ) to be less than a predefined value. By doing so, the optimization problem can be obtained as

$$\begin{aligned} \min_{\mathscr{Q}} \quad & trace[(U + B_d^\top S B_d)\mathscr{Q}] \\ \text{subject to} \quad & 2trace(C^\top \mathscr{P}^{-1} C \mathscr{U})\mathscr{T} \geq \Gamma \\ & \mathscr{U} - B_d \mathscr{Q} B_d^\top = \mathscr{A}\mathscr{U}\mathscr{A}^\top, \end{aligned} \tag{42}$$

where Γ is the lower bound of the expected value of Δg_k. As shown in [20], the solutions of two optimization problems (41) and (42) are scalar multiples of each other. Therefore, the solving either optimization problem guarantees same performance.

5. Simulation results

In this section, the theoretical results is evaluated through intensive simulation studies carried out to detect a replay attack in a quadrotor. First, we will represent the mathematical model of a quadrotor. Then, the watermarking framework will be applied to this system.

Dynamical behavior of a quadrotor can be modeled by nonlinear differential equations [3]. Since the watermarking framework is developed for a system in steady-state, we assume that the quadrotor is hovering around an equilibrium point. Then, we can linearize its nonlinear model around this point. After linearization, the system can be represented in state space form of (1). Matrices A and B are presented in [3]. The state variables vector is

$$x = [\dot{p}_x \quad p_x \quad \dot{p}_y \quad p_y \quad \dot{p}_z \quad p_z \quad \dot{\phi} \quad \phi \quad \dot{\theta} \quad \theta \quad \dot{\psi} \quad \psi]^\top, \tag{43}$$

where p_x, p_y, and p_z are used to determine position of the quadrotor in three principal directions; and ϕ, θ, and ψ are the roll, pitch and yaw angles, respectively; The output of the system is defined as $y = [p_x \quad p_y \quad p_z \quad \psi]^\top$; and Control input is defined as follows

$$u = \begin{bmatrix} F & \tau_\phi & \tau_\theta & \tau_\psi \end{bmatrix}^\top, \tag{44}$$

where F is the force that acts on the quadrotor, and τ_ϕ, τ_θ, and τ_ψ are rolling, pitching, and yawing torques. Physical parameters of the quadrotor including mass and moments of inertia are set to $m = 0.6\text{kg}$, $J_x = J_y = 0.0092\text{kgm}^2$, and $J_z = 0.0101\text{kgm}^2$. The process and measurement noises are considered as independent Gaussian random variables with zero mean.

Fig. 1 shows the expectation of Δg_k for different sampling periods. The optimization problem (41) is solved by using CVX [9]. It can be observed that when the sampling period is very small, $\mathbb{E}[\Delta g_k]$ becomes very small, then by increasing the sampling period, $\mathbb{E}[\Delta g_k]$ increases, but after $T = 0.1$, it starts decreasing. Hence, the optimal value of the sampling period for our system is $T = 0.1\text{s}$. The covariance matrix of the authentication signal for $T = 0.1\text{s}$ is obtained as

$$\mathscr{Q} = \begin{bmatrix} 1.4555 & 0.0207 & -0.0191 & 0.0911 \\ 0.0207 & 0.2003 & -0.0003 & 0.0298 \\ -0.0191 & -0.0003 & 0.2003 & -0.0274 \\ 0.0911 & 0.0298 & -0.0274 & 3.0043 \end{bmatrix} \tag{45}$$

To show the effect of sampling period on the detector's performance, the watermarking framework is applied to the quadrotor with different sampling periods. Simulation results are shown in Fig. 2. This figure shows several ROC curves for the quadrotor system under a replay attack. Six different sampling periods are used,

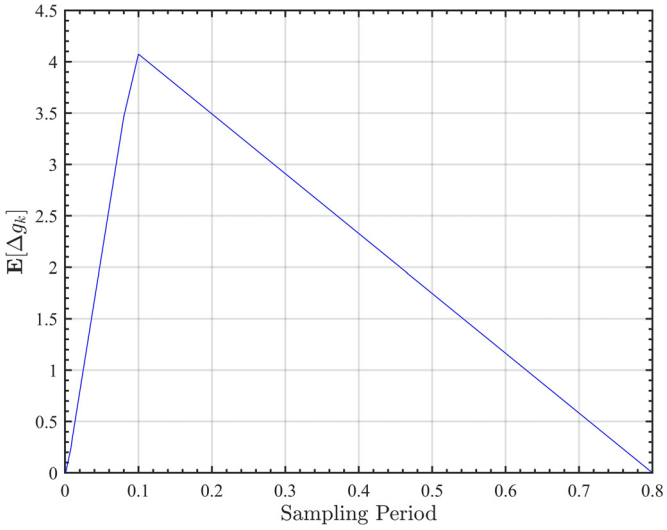

Fig. 1. Expectation of Δg_k for different sampling periods.

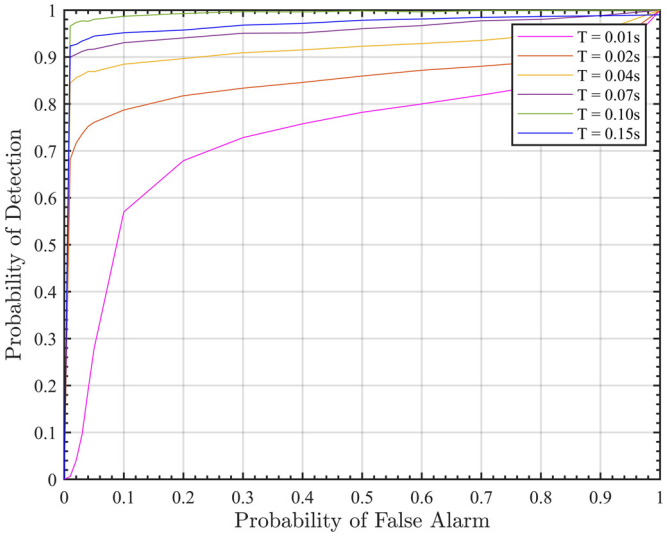

Fig. 2. ROC curve for replay attack detection in the quadrotor system with different sampling periods.

Table 1
LQG cost function for different values of sampling period.

T	0.02	0.04	0.07	0.10	0.15
J_T/J_1	1.03	1.10	1.21	1.38	1.68

and it can be observed that by increasing the sampling period, performance of the detector improves. However, after $T = 0.1\text{s}$ the ROC curve starts to change in the reverse direction and detector performance degrades. This behavior of the detector was expected, because the maximum value of $\mathbb{E}[\Delta g_k]$ occurs in $T = 0.1\text{s}$. To show the effect of sampling period on the LQG performance, cost function for these sampling periods are calculated and the lowest sampling period, $T = 0.01\text{s}$ is considered as the reference cost function (J_1) and the ratio of the cost function for other sampling periods (J_T) to this reference value is shown in Table 1.

6. Conclusion

In this paper, we generalize the design of physical watermarking to detect replay attack on digitally controlled continuous-time systems. In particular, we investigate the effect of sampling pe-

riod on the performance of the detector. We show that an optimal sampling period exists and we generalize the optimal watermarking signal design to include sampling period as a design variable jointly with the covariance of the watermark. Finally, we apply the watermarking framework to a quadrotor, and numerical simulations are included to illustrate our findings and validate our design.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the National Science Foundation, under grant number 1932530.

References

[1] K.J. Åström, Introduction to Stochastic Control Theory, Courier Corporation, 2012.
[2] K.J. Astrom, Introduction to Stochastic Control Theory, Elsevier, 1971.
[3] R.W. Beard, Quadrotor Dynamics and Control Rev 0.1 19 (3) (2008) 46–56.
[4] D.U. Case, Analysis of the cyber attack on the ukrainian power grid, Electr. Inf. Shar. Anal. Center (E-ISAC) (2016).
[5] G.F. Franklin, J.D. Powell, M.L. Workman, et al., Digital Control of Dynamic Systems, 3, Addison-wesley Reading, MA, 1998.
[6] M. Ghaderi, K. Gheitasi, W. Lucia, A blended active detection strategy for false data injection attacks in cyber-physical systems, IEEE Trans. Control Netw. Syst. 8 (1) (2020) 168–176 https://ieeexplore.ieee.org/document/9199094.
[7] A. Gelb, Applied Optimal Estimation, MIT Press, 1974.
[8] P. Griffioen, S. Weerakkody, B. Sinopoli, O. Ozel, Y. Mo, A tutorial on detecting security attacks on cyber-physical systems, in: Proceedings of the 18th European Control Conference (ECC), IEEE, 2019, pp. 979–984.
[9] M. Grant, S. Boyd, in: CVX: matlab software for disciplined convex programming, version 2.1, 2014.
[10] M. Hosseinzadeh, B. Sinopoli, E. Garone, Feasibility and detection of replay attack in networked constrained cyber-physical systems, in: Proceedings of the 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2019, pp. 712–717.
[11] S.H. Kafash, N. Hashemi, C. Murguia, J. Ruths, Constraining attackers and enabling operators via actuation limits, in: Proceedings of the IEEE Conference on Decision and Control (CDC), IEEE, 2018, pp. 4535–4540.
[12] M.J. Khojasteh, A. Khina, M. Franceschetti, T. Javidi, Learning-based attacks in cyber-physical systems, IEEE Trans. Control Netw. Syst. (2020).
[13] M.J. Khojasteh, A. Khina, M. Franceschetti, T. Javidi, Authentication of cyber-physical systems under learning-based attacks, IFAC-PapersOnLine 52 (20) (2019) 369–374.
[14] R. Langner, Stuxnet: dissecting a cyberwarfare weapon, IEEE Secur. Privacy 9 (3) (2011) 49–51.
[15] E.A. Lee, Cyber physical systems: design challenges, in: Proceedings of the 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC), IEEE, 2008, pp. 363–369.
[16] J. Markoff, A Silent Attack, But not a Subtle One, New York Times 160(55176) (2010) 6.
[17] R.K. Mehra, J. Peschon, An innovations approach to fault detection and diagnosis in dynamic systems, Automatica 7 (5) (1971) 637–640.
[18] Y. Mo, B. Sinopoli, Secure control against replay attacks, in: Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2009, pp. 911–918.
[19] Y. Mo, T.H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, B. Sinopoli, Cyber–physical security of a smart grid infrastructure, Proc. IEEE 100 (1) (2011) 195–209.
[20] Y. Mo, R. Chabukswar, B. Sinopoli, Detecting integrity attacks on SCADA systems, IEEE Trans. Control Syst. Technol. 22 (4) (2013) 1396–1407.
[21] D. Mustafa, D.S. Bernstein, LQG cost bounds in discrete-time H2/H-infinity control, Trans. Inst. Meas. Control 13 (5) (1991) 269–275.
[22] V. Renganathan, N. Hashemi, J. Ruths, T.H. Summers, Distributionally robust tuning of anomaly detectors in cyber-physical systems with stealthy attacks, in: Proceedings of the American Control Conference (ACC), IEEE, 2020, pp. 1247–1252.
[23] R. Romagnoli, S. Weerakkody, B. Sinopoli, A model inversion based watermark for replay attack detection with output tracking, in: Proceedings of the American Control Conference (ACC), IEEE, 2019, pp. 384–390.
[24] M.A. Rotea, The generalized H2 control problem, Automatica 29 (2) (1993) 373–385.
[25] D.E. Sanger, Obama Order Sped Up Wave of Cyberattacks Against Iran, 1, The New York Times, 2012, p. 2012.