Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Covert Attack Detection Based on $\mathcal{H}_i/\mathcal{H}_\infty$ Optimization for Cyber-Physical Systems ⋆

Jiao Qin * **Maiying Zhong** * **Yang Liu** * **Xianghua Wang** *
Donghua Zhou *,**

* *College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao 266590, China(e-mail: myzhong@buaa.edu.cn*
** *Department of Automation, Tsinghua University, Beijing 100084, China*

Abstract: In this paper, a new detection scheme is proposed to detect covert attack in the framework of $\mathcal{H}_i/\mathcal{H}_\infty$ index optimization for cyber-physical system (CPS) which is modeled as a linear discrete time-varying (LDTV) system. First, a random modulation matrix that the attacker cannot know is inserted into the path of the control variables to destroy the stealthiness of covert attacks. Second, a detection filter is constructed which transforms the detection problem into an $\mathcal{H}_-/\mathcal{H}_\infty$ or $\mathcal{H}_\infty/\mathcal{H}_\infty$ index optimization problem. The optimal solution is obtained by solving the Riccati equation. Third, a decision making mechanism is presented to trigger an alarm and further determine whether the cause of alarm is a covert attack or a fault. Finally, a simulation example is given to illustrate the effectiveness of the proposed method.

Keywords: Cyber-physical system, covert attack, LDTV system, detection filter, $\mathcal{H}_i/\mathcal{H}_\infty$ optimization.

1. INTRODUCTION

As a multi-dimensional complicated system that conforms information network and physical environment, CPS has become a hot research topic with the development of network and information technology in recent years. Through the organic integration and deep collaboration of computing, communication and control technologies, information space can be used to remotely manipulate a physical entity to realize actual-time perception and dynamic control of large engineering systems(Liu et al. (2019); Alguliyev et al. (2018)). Therefore, the security of both information and physical space needs to be taken into consideration.

Cyber attack is a natural development on the basis of physical attack with the emergence of information network. It is not limited by distance, easier to copy, and less dangerous for attackers to be discovered(Finogeev et al. (2017)). In CPS, the deep combination of information space and physical space produces obvious scientific superiority, but also makes it possible for attackers to invade the physical space by attacking the information space atthe same time. A famous example is an attack on an Australia sewage control system in 2000(Slay et al. (2007)). In the four months after being attacked, the pump could not operate when needed and communication between the control center and pumping station was disconnected which directly caused flooding of sewage near the factory. Another example for a replay attack is the Stuxnet virus discovered in 2010(Collins et al. (2012)). Attackers could attack the programmable logic controller in the industrial control system to modify the original program on computer according to their own wishes. Therefore, the research of cyber attack detection in CPS has received extensive attention in the late years. Cárdenas et al. (2008) pointed out that some characteristics of CPS make its security more challenging than internet technology systems, and a new mathematical framework for studying CPS attacks was given. From the perspective of attackers, Teixeira et al. (2015) defined an attack space by the adversary's model information, exposure, and interruption resources. According to existing research, cyber attacks can be categorized into denial of service (DoS) attacks and deception attacks which include replay attacks, covert attacks and zero-dynamics attacks. Since the DoS attacks directly block the signal transmission between controller and sensor, it is easy to be detected. However, deception attacks can avoid the general detection by obtaining the model knowledge and the reading authority of data transmission channel. Mo et al. (2009) defined a replay attack model and analyzed its effects on a control system. The idea to modify the system input behavior to disclose covert attacks was first proposed in Teixeira et al. (2012). With the aim of detecting replay attacks, Hoehn et al. (2016,a) inserted a nonlinear component in the control loop to stimulate the system in non-regular time pauses.

Cyber attack is similar to fault in some aspects, so there are some research results that apply fault diagnosis method to cyber attack detection in recent years. Keller et al. (2013) used the Kalman filter to detect zero-dynamics attacks. Lv et al. (2019) created an integral sliding mode observer to oversee cyber attacks. An Krein space-based method was employed to detect deception attacks for a dis-

⋆ This work was supported in part by the National Natural Science Foundation of China (Grants Nos. 61873149, 61733009, 61703244), and the Research Fund for the Taishan Scholar Project of Shandong Province of China.

crete time-varying system monitored by a sensor network in Ge et al. (2019).

As a representative model-based fault diagnosis technique, some works formulated the fault detection filter(FDF) design problems in the framework of $\mathcal{H}_-/\mathcal{H}_\infty$ or $\mathcal{H}_\infty/\mathcal{H}_\infty$ filtering(Zhong et al. (2018)). Li et al. (2009) demonstrated that the robust fault detection problems under various performance indices $\mathcal{H}_-/\mathcal{H}_\infty$, $\mathcal{H}_\infty/\mathcal{H}_\infty$ and $\mathcal{H}_2/\mathcal{H}_\infty$ can be worked out by an integrated optimal solution. In Zhong et al. (2010), the unified solution was utilized to FDF design of LDTV systems and can be acquired by solving the discrete time Riccati equation. Hoehn et al. (2016,b) proposed a method of inserting the modulation matrix in the control loop and using $\mathcal{H}_-/\mathcal{H}_\infty$ filter to detect covert attacks and zero-dynamic attacks. However, since Hoehn et al. (2016,b) only targeted linear time-invariant (LTI) system, the modulation matrix was constant or periodic, which was still easy to be obtained by attackers.

The innovation of this paper is reflected in the following three aspects. 1) This paper considers the problem of covert attack detection for CPSs whose plants are LDTV systems, which increases the design complexity. To destroy the stealthiness of attacks, a random modulation matrix that the attacker cannot know is inserted into the path of the control variables. Compared with constant or periodic matrix used in Hoehn et al. (2016,b), it is more difficult to be got by attackers. 2) We proposes an algorithm using the detection filter and $\mathcal{H}_i/\mathcal{H}_\infty$ index optimization to detect the covert attack while avoiding the excessive computation. The optimal solution is obtained by solving the Riccati equation. 3) A decision making mechanism is constructed to detect covert attack.We also consider the occurrence of faults and further determine whether the cause of alarm is a covert attack or a fault by changing the value of modulation matrix.

This paper is organized as follows. Problem formulation and the necessary preliminaries are recommended in Section 2, the basic ideas and main results will be introduced in Section 3. To illustrate this approach, a simulation example is given in Section 4. Finally, in Section 5 conclusions are drawn.

Notation. The notations used here are fairly standard. The superscripts '-1' and 'T' respectively on behalf of the inverse and transpose of a matrix. R^n denotes the n dimensional Euclidean space. The notation $X > 0$ (respectively, $X < 0$) means that X is a real positive definite (respectively, negative definite) matrix. $\|M\|$, $\|M\|_2$ and $\|M\|_-$ respectively refer to the Euclidean norm, the 2-norm and smallest singular value of matrix M. $\theta(k) \in l_2[0, N]$ means $\sum_{k=0}^{N} \theta^T(k)\theta(k) < \infty$. The symbol I denotes the identity matrix with appropriate dimension.

2. PRELIMINARIES AND PROBLEM FORMULATION

2.1 Covert attack of CPS

In a typical situation, a CPS generally consists of the remote plant, the transmission network and the local controller(Alguliyev et al. (2018)). The plant is represented as an LDTV system

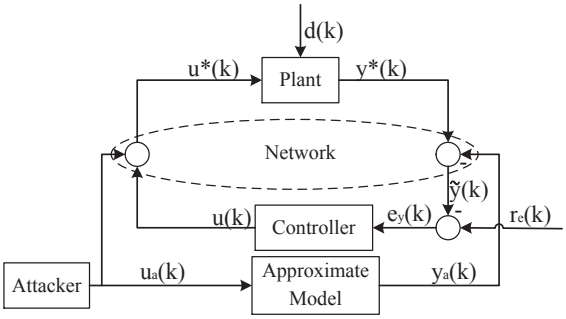

Fig. 1. Covert attack in CPS

$$\begin{cases} x(k+1) = A(k)x(k) + B(k)u(k) + B_d(k)d(k) \\ y(k) = C(k)x(k) + D_d(k)d(k) \end{cases},$$

where $x(k) \in R^n$, $u(k) \in R^q$, $y(k) \in R^m$, $d(k) \in l_2[0, N]$ are the state, control input, measured output vector, and unknown system noise respectively. $A(k)$, $B(k)$, $C(k)$, $B_d(k)$, and $D_d(k)$ are known time-varying matrices with suitable dimensions. For the sake of ensuring the existence of exponentially stable detection filter, it is assumed that$(C(k), A(k))$ is uniformly detectable and $(A(k), B_d(k))$ is uniformly stabilizable(Engwerda (1990)).

In practical applications, the controller should be robust to some common phenomena in the network. Similarly to Teixeira et al. (2015), the coding and decoding problems, delays, packet loss, etc. are not considered here. After observing the signal channels and obtaining enough system information, the attacker can identify an approximate model of the plant as follows:

$$\begin{cases} x_a(k+1) = \bar{A}(k)x_a(k) + \bar{B}(k)u_a(k) \\ y_a(k) = \bar{C}(k)x_a(k) \end{cases},$$

where $u_a(k)$ is the attack signal generated by the attacker, and $y_a(k)$ is the output of the approximate model, which indicates the attack influence. Without loss of generality, assume that $u_a(k) \in l_2[0, N]$. $\bar{A}(k)$, $\bar{B}(k)$ and $\bar{C}(k)$ denote, respectively, approximate matrices constructed by the attacker of $A(k)$, $B(k)$ and $C(k)$.

The attacker can change the actuator signals by accessing $u_a(k)$ to the inputs and $y_a(k)$ to the outputs. As shown in Fig. 1, when the CPS is attacked by covert attack, the control input forced on the plant is not $u(k)$ but $u^*(k)$, where $u^*(k) = u(k)+u_a(k)$. Similarly, the measured output after network transmission is not $y(k)$ but $\tilde{y}(k)$, where $\tilde{y}(k) = y^*(k) - y_a(k)$, $y^*(k)$ is the real output of the attacked plant. $r_e(k)$ is the reference signal and $e_y(k) = r_e(k) - y(k)$ is the output tracking error which is also the input of the controller. Assume that the approximate model established by the attacker is sufficiently accurate, which means $\bar{A}(k) = A(k)$, $\bar{B}(k) = B(k)$ and $\bar{C}(k) = C(k)$, and let $x^*(k)$ denote the state of the system after being attacked, then the plant dynamics under the covert attack is described by

$$\begin{cases} x^*(k+1) = A(k)x^*(k) + B(k)u^*(k) + B_d(k)d(k) \\ \tilde{y}(k) = C(k)x^*(k) + D_d(k)d(k) - y_a(k) \end{cases}.$$

According to the superposition principle of linear system, we have $x^*(k) = x(k) + x_a(k)$ and $u^*(k) = u(k) + u_a(k)$,

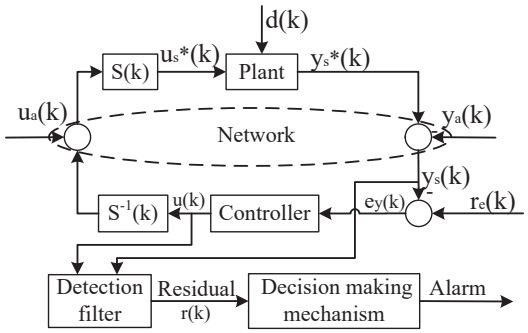

Fig. 2. Detection of covert attacks

hence $\tilde{y}(k) = y^*(k) - y_a(k) = y(k)$. That means the influence of attack signal $u_a(k)$ can be completely offset by subtracting $y_a(k)$ from the measurement signal $y^*(k)$, i.e., the attack is invisible. However, the intrusion of covert attacks makes the state of the system changes, which has a devastating effect on the normal operation of the system.

2.2 Problem formulation

Because the covert attack is invisible, that is, the input and output of the CPS do not contain the information of covert attack, it can not be detected by conventional methods. Inspired by the modulation matrix in Hoehn et al. (2016,b), we will insert random matrices $S(k)$ and $S^{-1}(k)$ into the path of the control variables, where $S(k) = diag\{s_i(k)\}, i = 1, 2, \ldots, n$. $s_i(k)$ is a constant randomly generated at time k, and the range can be artificially given. As shown in Fig. 2, $y_s^*(k)$ is the real output of the plant while the measured output of CPS is $y_s(k)$ which is transmitted to the controller. Then the system dynamics is given by

$$\begin{cases} x_s^*(k+1) = A(k)x_s^*(k) + B(k)u(k) + B_d(k)d(k) \\ \qquad\qquad + B(k)S(k)u_a(k) \\ y_s(k) = C(k)x_s^*(k) + D_d(k)d(k) - y_a(k) \end{cases}. \quad (1)$$

Remark 1. Both the modulation matrix $S(k)$ and $S^{-1}(k)$ are inserted after the attacker obtaining enough system information. In practical applications, we hope that the modulation matrix is easy to implement and difficult to be obtained by the attacker. Thereby we can destroy the information integrity and make the covert attack appear.

On the basis of the superposition principle, in view of $x_s^*(k) = x_s(k) + x_a(k)$, we have:

$$\begin{cases} x_s(k+1) = A(k)x_s(k) + B(k)u(k) + B_d(k)d(k) \\ \qquad\qquad + B(k)[S(k) - I]u_a(k) \\ x_a(k+1) = A(k)x_a(k) + B(k)u_a(k) \\ y_s(k) = C(k)[x_s(k) + x_a(k)] + D_d(k)d(k) \\ \qquad\qquad - C(k)x_a(k) \end{cases},$$

then the system can be obtained as follows:

$$\begin{cases} x_s(k+1) = A(k)x_s(k) + B(k)u(k) + B_d(k)d(k) \\ \qquad\qquad + B_a(k)u_a(k) \\ y_s(k) = C(k)x_s(k) + D_d(k)d(k). \end{cases}, \quad (2)$$

where $B_a(k) = B(k)[S(k) - I]$.

It is worth noting that the covert attacks $u_a(k)$ in CPS can be regarded as external intrusion signals such as faults, but there are still inherently distinct characteristics between them. On the one hand, faults are regarded as physical events that effect the behavior of system, where multiple faults occurring at the same time generally do not have a synergistic relationship. Covert attacks, however, may be simultaneously executed at multiple points in a coordinated way. On the other hand, faults generally occur randomly on system components, sensors, actuators or transmission channels, as opposed to covert attacks that do have a malicious intent(Keller et al. (2013); Rhouma et al. (2015)). In view of the difference between covert attack and fault, it is meaningful to improve and apply the fault detection methods to covert attack detection.

Based on the above preparations, the basic idea of this paper is to construct a covert attack detection filter using FDF design method, so that the problem can be transformed into an $\mathcal{H}_i/\mathcal{H}_\infty$ index optimization problem, and then get the filter parameters by solving Riccati equations. The specific content will be introduced separately in the following part.

3. MAIN RESULTS

3.1 Design of detection filter

Similar to FDF, the central mission of covert attack detection filter is to create a residual generation which is usually based on an observer. In this paper, the detection filter for system (2) can be constructed by

$$\begin{cases} \hat{x}_s(k+1) = A(k)\hat{x}_s(k) + B(k)u(k) \\ \qquad\qquad + L(k)[y_s(k) - \hat{y}_s(k)] \\ \hat{y}_s(k) = C(k)\hat{x}_s(k), \quad \hat{x}_s(0) = \hat{x}_{s0} \\ r(k) = V(k)[y_s(k) - \hat{y}_s(k)] \end{cases}, \quad (3)$$

where $\hat{x}_s(k) \in R^n$ is the state estimation vector of $x_s(k)$, $\hat{x}_{s0}$ is a guess of initial state, $r(k)$ is the residual, $L(k) \in R^{n\times m}$ and $V(k) \in R^{n\times m}$ are respectively observer gain matrix and post-filter that need to be designed. Define $e_x(k) = x_s(k) - \hat{x}_s(k)$, then the error dynamics can be obtained (by subtracting $\hat{y}_s(k)$ from $y_s(k)$) as

$$\begin{cases} e_x(k+1) = [A(k) - L(k)C(k)]e_x(k) + [B_d(k) \\ \qquad\qquad - L(k)D_d(k)]d(k) + B_a(k)u_a(k) \\ \varepsilon(k) = y_s(k) - \hat{y}_s(k) = C(k)e_x(k) + D_d(k)d(k) \\ r(k) = V(k)\varepsilon(k). \end{cases}.$$

Let

$$\begin{aligned} d_k &= [e_x^T(0), d^T(0), \ldots, d^T(k)]^T, \\ a_k &= [u_a^T(0), \ldots, u_a^T(k)]^T, \\ \mathrm{g}_{\varepsilon d}(k) &= [\mathrm{g}_e(k,0), \mathrm{g}_d(k,0), \ldots, \mathrm{g}_d(k,k-1), D_d(k)], \\ \mathrm{g}_{\varepsilon a}(k) &= [\mathrm{g}_a(k,0), \ldots, \mathrm{g}_a(k,k-1), \theta], \\ \Phi(k,i) &= \prod_{j=i}^{k-1}[A(j) - L(j)C(j)], \Phi(k,k) = I, \\ \mathrm{g}_e(k,0) &= C(k)\Phi(k,0), \\ \mathrm{g}_d(k,i) &= C(k)\Phi(k,i+1)[B_d(i) - L(i)D_d(i)], \\ \mathrm{g}_a(k,i) &= C(k)\Phi(k,i+1)B_a(i), i \le k-1, \end{aligned}$$

where θ is a zero matrix of m rows and q columns. Same as the FDF mentioned in Zhong et al. (2010), the residual generator can take the form as

$$r(k) = V(k)\mathrm{g}_{\varepsilon d}(k)d_k + V(k)\mathrm{g}_{\varepsilon a}(k)a_k.$$

The following performance indices, for $\forall k \in \mathcal{N}$, where $\mathcal{N}$ is denoted by a set with integers $0, 1, \ldots, N$, are defined as

$$\|G_{rd}\|_{\infty,[0,k]} = \sup_{d(k)\in l_2[0,N]} \frac{\sum_{i=0}^{k} \|V(i)\mathrm{g}_{\varepsilon d}(i)d_i\|^2}{\|e_x(0)\|^2 + \sum_{i=0}^{k} \|d(i)\|^2},$$

$$\|G_{ra}\|_{\infty,[0,k]} = \sup_{u_a(k)\in l_2[0,N]} \frac{\sum_{i=0}^{k} \|V(i)\mathrm{g}_{\varepsilon a}(i)a_i\|^2}{\sum_{i=0}^{k} \|u_a(i)\|^2},$$

$$\|G_{ra}\|_{-,[0,k]} = \inf_{u_a(k)\in l_2[0,N]} \frac{\sum_{i=0}^{k} \|V(i)\mathrm{g}_{\varepsilon a}(i)a_i\|^2}{\sum_{i=0}^{k} \|u_a(i)\|^2},$$

where $\|G_{rd}\|_{\infty,[0,k]}$ is used to evaluate the robustness of residual to noise, while the sensitivity of residual to covert attack is evaluated by $\|G_{ra}\|_{\infty,[0,k]}$ or $\|G_{ra}\|_{-,[0,k]}$, which represent the best and worst case sensitivity criteria, respectively. In order to detect covert attack effectively, we hope to decrease the affect of the noise $d(k)$ on the residual, simultaneously, intensity the sensitivity of the residual to the covert attack $u_a(k)$. So, the following maximization problem

$$\max_{L(k),V(k)} \frac{\|G_{rf}\|_{\infty,[0,k]}}{\|G_{rd}\|_{\infty,[0,k]}} \; or \; \max_{L(k),V(k)} \frac{\|G_{rf}\|_{-,[0,k]}}{\|G_{rd}\|_{\infty,[0,k]}} \tag{4}$$

can be used as an objective of covert attack detection filter design.

For a given system (2), the problem of designing filter (3) can be transformed into solving the optimal matrices $L_o(k)$ and $V_o(k)$ which make the system asymptotic stable and satisfy the performance specifications shown in (4) at each moment k. This optimization problem can be solved by the following Lemma 1.

Lemma 1. (Zhong et al. (2010)) Assume that $(C(k), A(k))$ is uniformly detectable and $(A(k), B_d(k))$ is uniformly stabilizable, then an optimal solution to both the finite horizon $\mathcal{H}_\infty/\mathcal{H}_\infty$ and $\mathcal{H}_-/\mathcal{H}_\infty$ detection filter problems is the $L_o(k)$ and $V_o(k)$ given by

$$L_o(k) = [A(k)P_o(k)C^T(k) + B_d(k)D_d^T(k)]R_d^{-1}(k) \tag{5}$$

$$V_o(k) = R_d^{-\frac{1}{2}}(k) \tag{6}$$

where $R_d(k) = C(k)P_o(k)C^T(k) + D_d(k)D_d^T(k) > 0$ and $P_o(k) \geq 0$ is the solution of the following Riccati equation

$$P_o(k+1) = A(k)P_o(k)A(k) - L_o(k)V_o^{-2}(k)L_o^T(k) + B_d(k)B_d^T(k), P_o(0) = I.$$

Remark 2. According to Zhong et al. (2010), $\|G_{rd}\|_{\infty,[0,k]}$ is still a reasonable index to appraise the robustness of residual to noise in case that $d(k)$ are stochastic noise sequences.

3.2 Design of decision making mechanism

If the plant is not affected by noise, i.e. $d(k) = 0$, then $r(k) = V(k)\mathrm{g}_{\varepsilon a}(k)a_k$. The threshold can be selected as 0 and the triggering alarm strategy can be set as follows:

$$\begin{cases} r(k) = 0 \Rightarrow no\ alarm \\ r(k) \neq 0 \Rightarrow alarm \end{cases}.$$

In the case where the system is disturbed by noise, set the time window size as N_s, and use the root mean square value of the signal $r(k)$ to define the residual evaluation function as follows (Ding. (2008)):

$$J_e(k) = \begin{cases} (\frac{1}{k}\sum_{i=1}^{k} \|r(k)\|^2)^{\frac{1}{2}}, k < N_s \\ (\frac{1}{N_s}\sum_{i=0}^{N_s-1} \|r(k-i)\|^2)^{\frac{1}{2}}, k \geq N_s \end{cases}. \tag{7}$$

The threshold is selected as

$$J_{th} = \sup_{u_a(k)=0} J_e(k) = \delta_e + \delta_d, \tag{8}$$

where $\delta_e \geq \|e_x(0)\|_2^2$, $\delta_d \geq \|d(k)\|_2^2$. The triggering alarm strategy can be set as follows:

$$\begin{cases} J_e(k) \leq J_{th} \Rightarrow no\ alarm \\ J_e(k) > J_{th} \Rightarrow alarm \end{cases}.$$

Then we can outline the attack detection procedure step by step.

Step 1: Choose a random modulation matrix $S(k)$ with a suitable variation range, and guarantee that $S(k) \neq I$.

Step 2: Calculate the optimal matrices $L_o(k), V_o(k)$ according to (5) (6), and then construct the filter (3) to generate a residual $r(k)$.

Step 3: Select the appropriate time window n_s and the constant ρ, then calculate the residual evaluation function $J_e(k)$ by (7), and compare it with the threshold J_{th} in (8) to determine whether to alarm.

It should be pointed out that when the system has fault, the residual $r(k)$ obtained by filter (3) will also be affected and alarm will be generated. Therefore, it is necessary to further determine whether the cause of the alarm is a covert attack or a fault. Without losing generality, it is assumed that covert attacks and faults do not occur at the same time. Based on above description, it can be easily known that a covert attack is detectable as far as the modulation matrix is not a unit matrix. In fact, when $S(k) = I$, the proposed detection filter becomes the normal FDF. Given a CPS which dynamics can be described by system (1), the detection filter (3) and the residual evaluation function (7) can be established. After the system generates an alarm, set $S(k) = I$, if the alarm signal disappears, it can be judged that the system has covert attack. Conversely, if the alarm signal still exists, it can be determined that the system exception is caused by fault. When the value of modulation matrix $S(k)$ can be changed artificially online, the proposed detection filter (3) can detect covert attack and separate it from fault.

4. AN ILLUSTRATIVE EXAMPLE

To verify the validity of the presented algorithm, conceive the system (2) with the following parameter matrices:

$$A(k) = \begin{bmatrix} 0.2 & 0 \\ -0.25 & 0.0015k \end{bmatrix}, B(k) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, C(k) = [\,-0.2\ \ 1\,],$$

$$B_d(k) = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}, D_d(k) = 2, B_a(k) = \begin{bmatrix} 0 \\ S(k) - 1 \end{bmatrix},$$

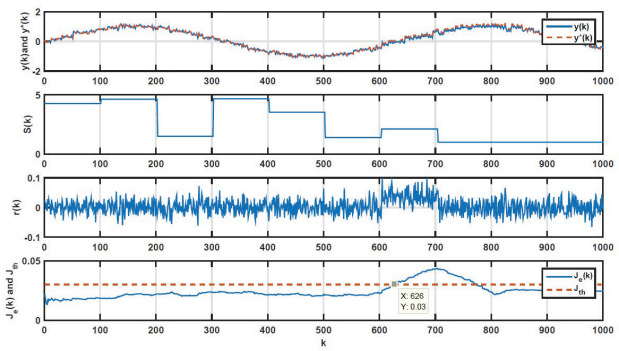

Fig. 3. CPS under covert attack.

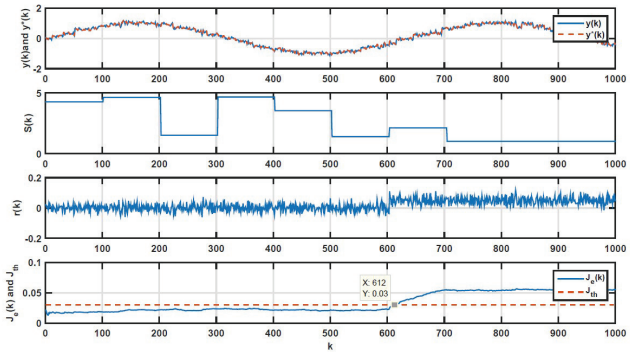

Fig. 4. CPS under sensor fault.

where $d(k)$ is a uniformly distributed noise bounded by $[0, 0.1]$. Set the total simulation time as $100s$ and the sampling period $T = 0.1s$, hence $k \in [0, 1000]$. A sinusoidal signal with amplitude of 1 and frequency of 0.1rad/s is selected as the reference input $r_e(k)$. In this example, The proportion-integration-differentiation(PID) incremental control algorithm used in this example is:

$$u(k) = u(k-1) + K_p[e_y(k) - e_y(k-1)] + K_p\frac{T}{T_i}e_y(k)$$
$$+K_p\frac{T_d}{T}[e_y(k) - 2e_y(k-1) + e_y(k-2)]$$

Adjust the parameters K_p, T_i, T_d to make the closed-loop system stable and the output tracking error $e_y(k) = \tilde{y}(k) - r_e(k)$ approaches zero. Add the modulation matrix $S(k)$, where $s_i(k) \in [1, 5]$ is randomly changed every 100 sampling periods, and the output signals can be shown in Fig. 3 and Fig. 4.

It can be seen from Fig. 3 that when the CPS is attacked by a covert attack signal $u_a(k)$ which is selected as the step function with a value of 0.1 starts at $60s(i.e., k = 600)$, the actual output $y^*(k)$ begins to deviate from the set value at $k = 601$. However, due to the presence of $y_a(k)$, the detectable output y(k) does not change, so the attack is invisible. Let $N_s = 100, J_{th} = 0.03$, the residual evaluation function $J_e(k)$ can be calculated, where the system can trigger the alarm when $k = 626$, that is, 2.6s after being attacked. Then set $S(k) = I$ at $k = 700$, after 72 sample times we have $J_e(k) < J_{th}(k)$, so the system is no longer alarm.

Suppose the CPS has a sensor fault $f(k)$, which is a step function with a value of 0.15 starts at $60s(i.e., k = 600)$, then the output signals of CPS and the residual obtained by the same $\mathcal{H}_i/\mathcal{H}_\infty$ filter is shown in Fig. 4. Use the same residual evaluation function $J_e(k)$ and the threshold value J_{th}, the system can trigger the alarm when $k = 612$, that is, 1.2s after the sensor fault happened. Then set $S(k) = I$ at $k = 700$, we can see the system is still alarm in Fig. 4. The above simulation results show that the proposed method in this paper can be effectively applied to the detection of covert attack in CPS and can further separate it from fault.

5. CONCLUSION

In this paper, for the CPS which is modeled as an LDTV system, the covert attacks are detected by constructing a detection filter. We first insert a random modulation matrix that the attacker cannot known into the control loop. Then construct the detection filter to transform the detection problem into $\mathcal{H}_\infty/\mathcal{H}_\infty$ or $\mathcal{H}_-/\mathcal{H}_\infty$ index optimization problem. The optimal solution is acquired by solving the Riccati equation, and the detection of covert attack is successfully realized while avoiding the excessive computation. This paper also discuss the influence of fault on covert attack detection by changing the modulation matrix. Finally the illustrative simulation results express that the presented method in this paper can be effectively applied to the CPS.

REFERENCES

A.A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd Conference on Hot Topics in Security*, San Jose, 2008.

A. Hoehn, and P. Zhang. Detection of replay attacks in cyber-physical systems. In *Proceedings of American Control Conference*, Boston, 290-295, 2016.

A. Hoehn, and P. Zhang. Detection of covert attacks and zero dynamics attacks in cyber-physical systems. In *Proceedings of American Control Conference*, Boston, 302-307, 2016.

A.G. Finogeev, and A.A. Finogeev. Information attacks and security in wireless sensor networks of industrial SCADA systems. *Journal of Industrial Information Integration*, Boston, 5:6-16, 2017.

A. Teixeira, I. Shames, H. Sandberg, and K.H. Johansson. A Secure Control Framework for Resource-Limited Adversaries. *Automatica*, 51:135–148, 2015.

A. Teixeira, I. Shames, H. Sandberg, and K.H. Johansson. Revealing stealthy attacks in control systems. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, Lllinois, 1806–1813, 2012.

J.C. Engwerda. Stabilizability and detectability of discrete-time time-varying systems. *IEEE Transactions on Automatic Control*, 35(4):425–429, 1990.

J. Liu, J. Tian, J. Wang, H. Wu, L. Sun, Y. Zhou, C. Shen, and X. Guan. Integrated security threats and defense of cyber-physical systems. *ACTA AUTOMATICA SINICA*, 45:5–24, 2019. (in Chinese)

J. Slay, and M. Miller. Lessons Learned from the Maroochy Water Breach. In *Proceedings of the International Conference on Critical Infrastructure Protection*, 73–82, 2007.

J.Y. Keller, and D. Sauter. Monitoring of stealthy attack in networked control systems. In *Proceedings of the 2013 Conference on Control and Fault-Tolerant Systems (SysTol)*, Nice, 462–467, 2013.

M. Lv, W. Yu, Y. Lv, J. Cao, and W. Huang. An integral sliding mode observer for CPS cyber security attack detection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, doi: 10.1063/1.5092637, 2019.

M. Zhong, S.X. Ding, and E.L. Ding. Optimal fault detection for linear discrete time-varying systems. *Automatica*, 46:1395–1400, 2010.

M. Zhong, S.X. Ding, and E.L. Ding. A survey on modelbased fault diagnosis for linear discrete time-varying systems. *Neurocomputing*, 306:51–60, 2018.

R. Alguliyev, Y. Imamverdiyev, and L. Sukhostat. Cyber-physical systems and their security issues. *Computers in Industry*, 100:213–223, 2018.

S. Collins, and S. McCombie. Stuxnet: the emergence of a new cyber weapon and its implications. *Intelligence and Counter Terrorism*, 7:80–91, 2012.

S.X. Ding. Model-based fault diagnosis techniques: design-schemes, algorithms and tools. *Springer*, pages 163-244, 2008.

T. Rhouma, J.Y. Keller, D. Sauter, K. Chabir, and M.N. Abdelkrim. Active GLR detector for resilient LQG controller in networked control systems. In *Proceedings of the IFAC SAFEPROCESS*, Paris, 754–759, 2015.

X. Ge, Q. Han, M. Zhong, and X. Zhang. Distributed Krein space-based attack detection over sensor networks under deception attacks. *Automatica*, doi:10.1016/j.automatica.2019.108557, 2019.

X. Li, and K. Zhou. A time domain approach to robust fault detection of linear time-varying systems. *Automatica*, 45:94–102, 2009.

Y. Mo, and B. Sinopoli. Secure control against replay attacks. In *Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing* , Lllinois, 911-918, 2009.