An Online Kullback-Leibler Divergence-Based Stealthy Attack against Cyber-Physical Systems

Qirui Zhang, Kun Liu, *Senior Member, IEEE*, André M. H. Teixeira, *Member, IEEE*, Yuzhe Li, *Member, IEEE*,
Senchun Chai, *Senior Member, IEEE*, and Yuanqing Xia, *Senior Member, IEEE*

***Abstract*—This article investigates the design of online stealthy attacks with the aim of moving the system's state to a desired target. Different from the design of offline attacks, which is only based on the system's model, to design the online attack, the attacker also estimates the system's state with the intercepted data at each instant and computes the optimal attack accordingly. To ensure stealthiness, the Kullback-Leibler divergence between the innovations with and without attacks at each instant should be smaller than a threshold. We show that the attacker should solve a convex optimization problem at each instant to compute the mean and covariance of the attack. The feasibility of the attack policy is also discussed. Furthermore, for the strictly stealthy case with zero threshold, the analytic expression of the unique optimal attack is given. Finally, a numerical example of the longitudinal flight control system is adopted to illustrate the effectiveness of the proposed attack.**

***Index Terms*—Security of the Cyber-physical systems (CPSs), Kullback-Leibler divergence (KLD), online stealthy attack**

I. Introduction

ADVANCES in communications boost the applications of cyber-physical system (CPSs), which integrate the communication, computation, and control technologies [1] [2]. However, the extensive employment of open communication networks makes CPSs vulnerable to various malicious attacks on the cyber layer, which can lead to remarkable economy losses. Hence, the security of the CPSs has drawn significant attention from the scientific community in recent years [3]–[6].

To guarantee the safety of CPSs, passive [7] and active [8] attack detectors are designed to judge whether the system is attacked. Once an attack is detected, one may design filters or controllers to mitigate its effect. However, some attackers with enough resources can carefully design attacks so that they can avoid being detected. Hence, to study the vulnerability of CPSs, researchers study designing stealthy attacks and evaluate how much destruction these attacks make.

Many works are devoted to the stealthy attacks against a specific type of detector, e.g., the χ^2 detector [9], the CUSUM detector [10], the MEWMA detector [11], the SPRT detector [12]. Furthermore, since systems may not just equip one attack detector, stealthy attacks with respect to arbitrary detectors attract much interest. The existence of undetectable attack is analyzed in [13]. Specifically, the zero-dynamic attack [14], which is designed according to the system's unstable poles and makes the output the same as the normal one, is one kind of such attack. In [15], the stability of the system under the undetectable attack is further studied. Moreover, the cases that the attacker is only able to attack at most a certain number of sensors and actuators are considered in [16] and [17], respectively.

A commonly adopted measure of the stealthiness for arbitrary detectors is the Kullback-Leibler divergence (KLD). Generally, the smaller the KLD between the innovations with and without attack is, the more stealthy the attack is [18]. With the KLD as the constraint, the optimal stealthy attacks are designed to degrade the system's estimate performance [19] and control performance [20]. In [18]–[20], it is required that the KLD between the sequences of all the innovations with and without attacks is smaller than a threshold. The constraint that the KLD between the innovations with and without attacks at each instant is used in [21], where an innovation-based attack is designed to maximize the trace of the covariance of the estimate error. The objective of reducing the linear quadratic Gaussian control performance is further considered in [22] and the boundedness and the approximation of the reachable set for the system under stealthy attack is studied in [23].

It should be pointed out that the attacks designed in [18]–[20], [22] are offline. That is, the attacker determines the distribution of the attack will be used in the future according to the system model at instant 0. In [21], although online attack is designed, it only considers the attack which modifies the filter's innovation with a specific linear function.

In this article, we aim to design an online attack to move the system's state to a desired target while letting the KLD between the innovations with and without attacks at each instant smaller than a threshold. The design of the optimal online stealthy attack is based on not only the system model, but also the input and output signals intercepted at each instant. The main contributions of this article are summarized as follows:

1) An algorithm, in which a convex optimization problem should be solved at each instant, is provided to compute the optimal online attack, and the algorithm is proved to be always recursively feasible.
2) In particular, when the threshold of the KLD equals to zero, which forces the attack to be strictly stealthy, it is shown that the optimal attack is unique and the analytical expression of the optimal strictly stealthy attack is given.

The rest of the article is organized as follows: Section II presents the problem formulation. In Section III, we design the optimal strictly stealthy attack and the optimal attack that is not strictly stealthy. Section IV presents a simulation to illustrate the effectiveness of the stealthy attack. Section V concludes this article.

Notations: Throughout this article, let $\mathbf{R}^n$ be the n-dimensional Euclidean space and I_n be the identity matrix of order n. The expectation of the stochastic variable x is denoted as $\mathbb{E}\{x\}$ and $x \sim \mathcal{N}(a, \Sigma)$ means the vector x satisfies a Gaussian distribution with mean a and covariance matrix Σ (when $\Sigma = 0$, it means

This work is supported by the National Natural Science Foundation of China (no. 61873034), the Swedish Research Council (no. 2018-04396), the Swedish Foundation for Strategic Research, the Joint Open Fundation of the State Key Laboratory of Synthetical Automation for Process Industries (no. 2021-KF-21-05), the Natural Science Foundation of Beijing Municipality (no. 4182057), and the Open Subject of Beijing Intelligent Logistics System Collaborative Innovation Center (no. BILSCIC-2019KF-13).(Corresponding author: Kun Liu).

Qirui Zhang, Kun Liu, Senchun Chai, and Yuanqing Xia are with the School of Automation, Beijing Institute of Technology, Beijing 100081, China. E-mail: qiruizhang@cumt.edu.cn; kunliubit@bit.edu.cn; chaisc97@163.com; xia_yuanqing@bit.edu.cn

André M. H. Teixeira is with the Department of Information Thechnology, Uppsala University, PO Box 337, SE-75105, Uppsala, Sweden. E-mail: andre.teixeira@it.uu.se

Yuzhe Li is with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China. E-mail: yuzheli@mail.neu.edu.cn

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3192201

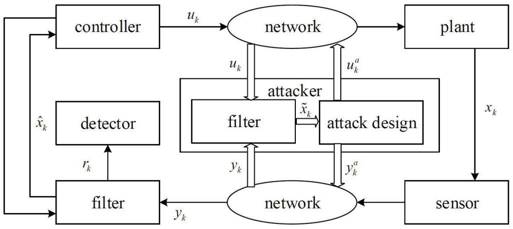

Fig. 1. Online stealthy deception attack design against CPSs

$x = a$). For any matrix A, $A^\dagger$ is the Moore-Penrose pseudoinverse, $\mathrm{Null}(A)$ stands for the null space of A, the column space of A is denoted by $\mathrm{Span}(A)$, and $\mathrm{Tr}(A)$ is used to denote the trace of A when A is a square matrix. The notation $\mathrm{diag}(A, B)$ stands for the block diagonal matrix with matrices A, B in the diagonal. For any symmetric matrix P, the notation $P \succ 0$ ($P \succeq 0$) means that P is positive definite (semi-definite) and $||x||_P^2 = x^T P x$, where x is a vector with appropriate dimension. For a sequence of vectors $x_i \in \mathbf{R}^n$, $i = a, a+1, \ldots, b$, the vector $x_a^b \in \mathbf{R}^{(b-a+1)n}$ equals to $\begin{bmatrix} x_a^T & x_{a+1}^T & \cdots & x_b^T \end{bmatrix}^T$.

II. Problem Formulation

In this section, we present the system model and describe the attacker's objective.

A. System model

Consider the system shown in Fig.1. The dynamics of the plant are given by

$$x_{k+1} = Ax_k + Bu_k + Du_k^a + w_k, \tag{1}$$

$$y_k = Cx_k + Ey_k^a + v_k, \tag{2}$$

where $x_k \in \mathbf{R}^n$ is the system state, $u_k \in \mathbf{R}^l$ is the control input, $u_k^a \in \mathbf{R}^p$ is the actuator attack, $y_k^a \in \mathbf{R}^q$ is the sensor attack, $w_k \in \mathbf{R}^n$ is the process noise, $y_k \in \mathbf{R}^m$ is the sensor output, and $v_k \in \mathbf{R}^m$ is the measurement noise. The process noise w_k has independent identical distribution (i.i.d.) $\mathcal{N}(0, \Sigma_w)$ with $\Sigma_w \succ 0$, the measurement noise v_k has i.i.d. $\mathcal{N}(0, \Sigma_v)$ with $\Sigma_v \succ 0$ and w_k is independent of v_k. The matrices A, B, C, D, and E have appropriate dimensions. Both D and E have full column rank. The pair (C, A) is observable, and the pair (A, B) is controllable. The system runs without knowing the attack signal. A Kalman filter is used to estimate the state. It is well-known that the Kalman filter converges exponentially fast from any initial condition [24]. Hence, without loss of generality we assume that the filter starts from the steady state, which makes the filter gain fixed, i.e., the Kalman filter has the form

$$\hat{x}_k = \hat{x}_k^- + K(y_k - C\hat{x}_k^-), \tag{3}$$

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}, \tag{4}$$

where $\hat{x}_k$ and $\hat{x}_k^-$ are the estimates of x_k given all the information up to instants k and $k-1$, respectively, $K = PC^T(CPC^T + \Sigma_v)^{-1}$ and $P = APA^T + \Sigma_w - APC^T(CPC^T + \Sigma_v)^{-1}CPA^T$. It is assumed that $A - KCA$ is stable.

Define the estimation error and the innovation of the Kalman filter as $e_k = x_k - \hat{x}_k$ and $r_k = y_k - C\hat{x}_k^-$, respectively. The attack detector uses r_k to judge whether the system is attacked according to a certain decision rule, e.g., [9]–[12].

Let $\bar{e}_k$ and $\bar{r}_k$ be the estimate error and the innovation of the Kalman filter when the system is under no attack (i.e., $u_{i-1}^a = 0$, $y_i^a = 0$, $i \leq k$), respectively, then $\bar{r}_k$ has i.i.d. $\mathcal{N}(0, S)$ with $S = CPC^T + \Sigma_v$.

In addition, the system uses a state feedback controller

$$u_k = L\hat{x}_k, \tag{5}$$

where L is the controller gain such that $A + BL$ is stable.

B. Stealthy attack

For the attacker, we assume that it has full knowledge of the system, i.e., the attacker knows the matrices A, B, C, D, E, L, Σ_w, and Σ_v. This assumption is the worst case for the system and is also necessary for the attacker to carefully design stealthy attacks. Hence, it is commonly used in the existing literature [18], [20]–[22]. Moreover, the attacker may acquire the parameters of the system from specific physical problems or by system identification techniques [25].

Let $\xi_k = \begin{bmatrix} (u_k^a)^T & (y_{k+1}^a)^T \end{bmatrix}^T$. Without loss of generality, the attack ξ_k is assumed to start at instant 0 and end at instant N. The attacker intercepts the system's input and output. The information obtained by the attacker at instant 0 is $\mathcal{I}_0 = \{u_{-\infty}^{-1}, y_{-\infty}^0\}$ and at instant $k \geq 1$ is $\mathcal{I}_k = \{u_{-\infty}^{k-1}, y_{-\infty}^k, \xi_0^{k-1}\}$. The attack policy is designed based on the information $\mathcal{I}_k$, i.e., $\xi_k = \mathcal{F}(\mathcal{I}_k)$. The attacker's goal is to find the optimal attack policy to make the system's state approach a certain target x^* while being stealthy.

With information $\mathcal{I}_k$, the attacker can give another estimate of the system, defined by $\tilde{x}_k$. The filter performed by the attacker has the following form:

$$\tilde{x}_k = \tilde{x}_k^- + K(y_k - Ey_k^a - C\tilde{x}_k^-), \tag{6}$$

$$\tilde{x}_k^- = A\tilde{x}_{k-1} + Bu_{k-1} + Du_{k-1}^a. \tag{7}$$

Since $\tilde{x}_k - \hat{x}_k$ converges to 0 as k approaches 0 from $-\infty$, we have $\tilde{x}_0 = \hat{x}_0$. Define the the innovation of filter (6) and (7) by $\tilde{r}_k = y_k - Ey_k^a - C\tilde{x}_k^-$. It can be easily proved that $\tilde{r}_k = \bar{r}_k$ for $k \geq 0$.

Next, we introduce the measure of stealthiness. The KLD, which is nonnegative and reflects the difference between two distributions, is widely used in the detection theory as mentioned in Section I. The definition of the KLD is given as follows:

Definition 1. *(KLD) [26] Let x and y be two random vectors with probability density functions f_x and f_y, respectively. The KLD between x and y is*

$$D(x||y) = \int_{\{\zeta | f_x(\zeta) > 0\}} f_x(\zeta) \log \frac{f_x(\zeta)}{f_y(\zeta)} d\zeta. \tag{8}$$

The KLD between filter's innovations with and without attacks from instants 1 to $N+1$, i.e., $D(r_1^{N+1}||\bar{r}_1^{N+1})$, is usually adopted to describe the stealthiness of the attack [18]–[20]. By the Chernoff-Stein Lemma [27], the value of $D(r_1^{N+1}||\bar{r}_1^{N+1})$ is related to the false alarm rate (i.e., the probability that the detector alarms but there is no attack) of any detector which uses the innovation r_1^{N+1} to judge whether there is an attack from instants 0 to N. However, most existing detectors (see [7] [8] and the references therein) judge whether the system is attacked at each instant. Hence, using $D(r_1^{N+1}||\bar{r}_1^{N+1})$ to describe the stealthiness of the attacks against these detectors is not suitable.

Following [21]–[23], we adopt $D(r_i||\bar{r}_i)$, $i = 1, \ldots, N+1$, as the measure of stealthiness. To be stealthy, the attacker should let $D(r_i||\bar{r}_i) \leq \delta_i$, where $\delta_i \geq 0$, $i = 1, \ldots, N+1$, are the thresholds.

Based on the above analysis, we consider the following optimization problem and find the optimal solution $\xi_k^* = \mathcal{F}^*(\mathcal{I}_k)$, $k = 0$, $\ldots, N$.

Problem 1.

$$\begin{aligned} \min_{\mathcal{F}(\mathcal{I}_0),\ldots,\mathcal{F}(\mathcal{I}_N)} \quad & J = \mathbb{E}\left\{\sum_{i=1}^{N+1} ||x_i - x^*||_{Q_i}^2\right\}, \\ \text{s.t.} \quad & D(r_i||\bar{r}_i) \le \delta_i, i = 1, \ldots, N+1, \end{aligned}$$

where $Q_i \succ 0$, $i = 1, \ldots, N+1$, are the weight matrices.

Remark 1. *The optimal offline attack policies are designed in [18]–[20], where $D(r_1^{N+1}||\bar{r}_1^{N+1})$ is used. By the chain rule of the KLD [27], we have $D(r_1^{k+1}||\bar{r}_1^{k+1}) = D(r_1||\bar{r}_1) + \sum_{i=2}^{k+1} \mathbb{E}\{D(r_i||\bar{r}_i|r_1^{i-1})\}$, where $D(r_i||\bar{r}_i|r_1^{i-1})$ is the conditional KLD between r_i and $\bar{r}_i$ given the condition r_1^{i-1}, and the expectation is taken over r_1^{i-1}. However, r_1^{i-1} becomes a known constant at instant i. Hence, using $D(r_1^{N+1}||\bar{r}_1^{N+1})$ is also not appropriate when we target to design an online attack policy, i.e., ξ_k is computed according to $\mathcal{I}_k$.*

III. Main Results

In this section, we will provide an algorithm that solves Problem 1 and further give the analytical solution for the case that the attack is strictly stealthy, i.e., $\delta_i = 0$, $i = 1, \ldots, N+1$.

Combining dynamics (1), (2), and filter (3), (4), we get

$$\begin{aligned} r_{k+1} &= CAe_k + G\xi_k + Cw_k + v_{k+1} \\ &= C(A\bar{e}_k + w_k) + v_{k+1} + CA\Delta e_k + G\xi_k \\ &= \bar{r}_{k+1} + CA\Delta e_k + G\xi_k, \end{aligned} \tag{9}$$

where $G = \begin{bmatrix} CD & E \end{bmatrix}$ and $\Delta e_k = e_k - \bar{e}_k$ is the estimate error induced by the attack.

Moreover, we also have

$$e_{k+1} = \bar{A}e_k + \bar{B}\xi_k + (I_n - KC)w_k - Kv_{k+1},$$

where $\bar{A} = A - KCA$ and $\bar{B} = \begin{bmatrix} D - KCD & -KE \end{bmatrix}$.

Hence, Δe_k satisfies the following equation:

$$\Delta e_{k+1} = \bar{A}\Delta e_k + \bar{B}\xi_k. \tag{10}$$

Since the attack starts at instant 0, we have $\Delta e_0 = 0$.

Let $z_k = \begin{bmatrix} \tilde{x}_k^T & \hat{x}_k^T & \Delta e_k^T & (x^*)^T \end{bmatrix}^T$. It follows from (3)-(7), (9), (10), and $\tilde{r}_k = \bar{r}_k$ that

$$z_{k+1} = \tilde{A}z_k + \tilde{B}\xi_k + H\tilde{r}_{k+1}, \tag{11}$$

where $\tilde{A} = \begin{bmatrix} A & BL & 0 & 0 \\ 0 & A+BL & KCA & 0 \\ 0 & 0 & \bar{A} & 0 \\ 0 & 0 & 0 & I_n \end{bmatrix}$, $H = \begin{bmatrix} K \\ K \\ 0 \\ 0 \end{bmatrix}$, and $\tilde{B} = \begin{bmatrix} D & 0 \\ KCD & KE \\ D-KCD & -KE \\ 0 & 0 \end{bmatrix}$.

Remark 2. *In (11), $\tilde{r}_{k+1} \sim \mathcal{N}(0, S)$ can be obtained from the attacker's filter* (6) *and* (7)*. The initial state $z_0 = \begin{bmatrix} \tilde{x}_0^T & \tilde{x}_0^T & 0 & (x^*)^T \end{bmatrix}^T$ is also known to the attacker. Hence, the attacker can run* (11) *with its own data, which means that z_k is fully accessible to the attacker.*

Define $F_1 = \begin{bmatrix} I_n & 0 & 0 & -I_n \end{bmatrix}$ and $F_2 = \begin{bmatrix} 0 & 0 & I_n & 0 \end{bmatrix}$, then we have $\tilde{x}_k - x^* = F_1 z_k$ and $\Delta e_k = F_2 z_k$. It is well-known that $x_k - \tilde{x}_k \sim \mathcal{N}(0, (I_n - KC)P)$ and is orthogonal to $\tilde{x}_k$ [24]. Therefore, we have

$$J = \mathbb{E}\left\{\sum_{i=1}^{N+1} \left[\text{Tr}((I_n - KC)PQ_i) + ||F_1 z_i||_{Q_i}^2\right]\right\}. \tag{12}$$

Then, the objective function in Problem 1 can be replaced by $\mathbb{E}\{\sum_{i=1}^{N+1} ||F_1 z_i||_{Q_i}^2\}$.

Following [18]–[23], we can prove that the optimal attack should have Gaussian distribution, i.e., $\xi_k \sim \mathcal{N}(\eta_k, \Gamma_k)$ with mean $\eta_k \in \mathbf{R}^{p+q}$ and covariance $\Gamma_k \succeq 0$, $k = 0, \ldots, N$. Hence, to obtain the optimal attack, we only need to calculate η_k and Γ_k.

A. Optimal Stealthy Attack

In this subsection, we will give the solution of Problem 1. To do so, we first define the following optimization problem:

Problem 2.

$$\begin{aligned} \min_{\Omega_k^N} \quad & \tilde{J}_k = \mathbb{E}_{z_{k+1}^{N+1}}\left\{\sum_{i=k+1}^{N+1} ||F_1 z_i||_{Q_i}^2 | \mathcal{I}_k\right\}, \\ \text{s.t.} \quad & D(r_i||\bar{r}_i|\mathcal{I}_k) \le \delta_i, i = k+1, \ldots, N+1, \end{aligned}$$

where $\Omega_k^N = \{\eta_k, \Gamma_k, \ldots, \eta_N, \Gamma_N\}$.

Problem 2 aims to find the optimal stealthy attack to minimize $\tilde{J}_k$ when $\mathcal{I}_k$ (or z_k, equivalently) is known. The relationship between Problems 1 and 2 will be discussed later.

Since the optimal attack is Gaussian, we shall express the constraint and the objective function of Problem 2 in terms of the mean $\eta_k, \ldots, \eta_N$, and the covariance $\Gamma_k, \ldots, \Gamma_N$ of the attack.

In (9), note that $\bar{r}_{k+1}$ is the innovation at instant $k+1$ when there is no attack, and the design of attack ξ_k is based on the information $\mathcal{I}_k$ at instant k. Hence, $\bar{r}_{k+1}$ is independent of ξ_k.

Given the condition $\mathcal{I}_k$, the term Δe_k is known. Then, according to (9) and (10), we have

$$r_i|\mathcal{I}_k \sim \mathcal{N}(\beta_{k,i}, S_{k,i}),$$

where

$$\begin{aligned} \beta_{k,i} &= CA\bar{A}^{i-k-1}\Delta e_k + \sum_{j=k}^{i-1} \bar{G}_{j,i}\eta_j, \\ S_{k,i} &= S + \sum_{j=k}^{i-1} \bar{G}_{j,i}\Gamma_j\bar{G}_{j,i}^T, \\ \bar{G}_{i-1,i} &= \ G, \\ \bar{G}_{j,i} &= CA\bar{A}^{i-2-j}\bar{B}, k \le j \le i-2. \end{aligned}$$

Recall that $\bar{r}_k$ is also Gaussian. Hence, it follows that

$$\begin{aligned} D(r_i||\bar{r}_i|\mathcal{I}_k) = \frac{1}{2}\Big[&\text{Tr}(S^{-1}S_{k,i}) + \log\frac{|S|}{|S_{k,i}|} - m \\ &+ \beta_{k,i}^T S^{-1}\beta_{k,i}\Big]. \end{aligned} \tag{13}$$

Let

$$\mathcal{A}_k = \begin{bmatrix} \tilde{A}^T & (\tilde{A}^2)^T & \cdots & (\tilde{A}^{N+1-k})^T \end{bmatrix}^T,$$

$$\mathcal{B}_k = \begin{bmatrix} \tilde{B} & 0 & \cdots & 0 \\ \tilde{A}\tilde{B} & \tilde{B} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{A}^{N-k}\tilde{B} & \tilde{A}^{N-1-k}\tilde{B} & \cdots & \tilde{B} \end{bmatrix},$$

$$\mathcal{H}_k = \begin{bmatrix} H & 0 & \cdots & 0 \\ \tilde{A}H & H & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{A}^{N-k}H & \tilde{A}^{N-1-k}H & \cdots & H \end{bmatrix},$$

then, from (11), one has

$$z_{k+1}^{N+1} = \mathcal{A}_k z_k + \mathcal{B}_k \xi_k^N + \mathcal{H}_k \tilde{r}_{k+1}^{N+1}.$$

Therefore, the objective function of Problem 2 can be written as

$$\begin{aligned}\tilde{J}_k =& \operatorname{Tr}(\tilde{Q}_k(\tilde{S}_k + \mathcal{A}_k z_k z_k^T \mathcal{A}_k^T + \mathcal{B}_k \tilde{\Gamma}_k \mathcal{B}_k^T)) \\ &+ 2 z_k^T \mathcal{A}_k^T \tilde{Q}_k \mathcal{B}_k \eta_k^N + ||\mathcal{B}_k \eta_k^N||_{\tilde{Q}_k}^2, \end{aligned} \quad (14)$$

where

$$\begin{aligned}\tilde{S}_k &= \mathcal{H}_k \operatorname{diag}(S, \ldots, S) \mathcal{H}_k^T, \\ \tilde{\Gamma}_k &= \operatorname{diag}(\Gamma_k, \ldots, \Gamma_N), \\ \tilde{Q}_k &= \operatorname{diag}(F_1^T Q_{k+1} F_1, \ldots, F_1^T Q_{N+1} F_1).\end{aligned}$$

In (13) and (14), note that the quadratic forms $\beta_{k,i}^T S^{-1} \beta_{k,i}$ and $||\mathcal{B}_k \eta_k^N||_{\tilde{Q}_k}^2$ as well as the function $-\log|S_{k,i}|$ are convex [28]. Moreover, other terms are linear functions of η_k, Γ_k, …, η_N, Γ_N. Hence, Problem 2 is convex and can be effectively solved by the CVX toolbox [29].

Suppose the solution of Problem 2 at instant k is $\{\eta_{k,k}, \Gamma_{k,k}, \ldots, \eta_{k,N}, \Gamma_{k,N}\}$. Then, we can present the algorithm that gives the optimal solution $\xi_k^* = \mathcal{F}^*(\mathcal{I}_k)$ of Problem 1 as follows:

Algorithm 1 Optimal attack policy $\mathcal{F}^*(\mathcal{I}_k)$

```
$\mathcal{I}_0 = \{u_{-\infty}^{-1}, y_{-\infty}^0\}$ and $k = 0$
while $0 \leq k \leq N$ do
    Calculate $z_k$ from $\mathcal{I}_k$ and solve Problem 2
    Attack the system with the optimal attack $\xi_k^* \sim \mathcal{N}(\eta_{k,k}, \Gamma_{k,k})$
    $k \leftarrow k + 1$
    $\mathcal{I}_k = \{\mathcal{I}_{k-1}, u_{k-1}, y_k, \xi_{k-1}\}$
end while
```

Next, we analyze the property of the covariances $\Gamma_{k,k}$, …, $\Gamma_{k,N}$ solved from Problem 2 at instant k, which will be used to prove that Algorithm 1 is always recursively feasible, i.e., at each instant, Problem 2 always has at least one solution.

Lemma 1. *The covariances* $\Gamma_{k,k}$, …, $\Gamma_{k,N}$ *satisfy*

$$\bar{G}_{j,i} \Gamma_{k,j} \bar{G}_{j,i}^T = 0, i = k+1, \ldots, N+1, j = k, \ldots, i-1, \quad (15)$$

$$\operatorname{Tr}(\tilde{Q}_k \mathcal{B}_k \operatorname{diag}(\Gamma_{k,k}, \ldots, \Gamma_{k,N}) \mathcal{B}_k^T) = 0. \quad (16)$$

Proof. In (13), note that $\operatorname{Tr}(S^{-1} S_{k,i}) + \log \frac{|S|}{|S_{k,i}|} \geq m$, where the equality holds if and only if $S_{k,i} = S$, which means (13) is minimized when $\sum_{j=k}^{i-1} \bar{G}_{j,i} \Gamma_j \bar{G}_{j,i}^T = 0$, $i = k+1$, …, $N+1$. Since $\bar{G}_{j,i} \Gamma_j \bar{G}_{j,i}^T \succeq 0$, we can obtain (15).

In (14), since $\tilde{Q}_k \succeq 0$, the term $\operatorname{Tr}(\tilde{Q}_k \mathcal{B}_k \tilde{\Gamma}_k \mathcal{B}_k^T)$ is minimized when (16) is satisfied. Note that (15) and (16) can be satisfied simultaneously when $\Gamma_{k,k} = \cdots = \Gamma_{k,N} = 0$. Hence, the proof is completed, □

Now, we are ready to study the feasibility of Algorithm 1.

Theorem 1. *Algorithm 1 is always recursively feasible.*

Proof. Note that $\Delta e_0 = 0$. Hence, Algorithm 1 is feasible at instant 0 with $\Omega_0^N = \{0, 0, \ldots, 0, 0\}$.

Suppose at instant $k-1$, Algorithm 1 is feasible and the system is attacked with the optimal attack $\xi_{k-1}^* \sim \mathcal{N}(\eta_{k-1,k-1}, \Gamma_{k-1,k-1})$. By (9), (10) and (15), for $i = k+1$, …, $N+1$, we have

$$r_i = \bar{r}_i + CA\bar{A}^{i-k-1} \Delta e_k + \sum_{j=k}^{i-1} \bar{G}_{j,i} \xi_j,$$

where

$$\begin{aligned} CA\bar{A}^{i-k-1} \Delta e_k &= CA\bar{A}^{i-k-1}(\bar{A} \Delta e_{k-1} + \bar{B} \xi_{k-1}^*) \\ &= CA\bar{A}^{i-k} \Delta e_{k-1} + \bar{G}_{k-1,i} \xi_{k-1}^* \\ &= CA\bar{A}^{i-k} \Delta e_{k-1} + \bar{G}_{k-1,i} \eta_{k-1,k-1}. \end{aligned}$$

Since $\{\eta_{k-1,k-1}, \Gamma_{k-1,k-1}, \ldots, \eta_{k-1,N}, \Gamma_{k-1,N}\}$ is the solution of Problem 2 at instant $k-1$, it follows from (15) that

$$\bar{G}_{j,i} \Gamma_{k-1,j} \bar{G}_{j,i}^T = 0, i = k, \ldots, N+1, j = k-1, \ldots, i-1. \quad (17)$$

In (13), replace k by $k-1$ and let $\eta_j = \eta_{k-1,j}$, $\Gamma_j = \Gamma_{k-1,j}$. Then, by (17), for $i = k$, …, $N+1$, one has

$$||CA\bar{A}^{i-k} \Delta e_{k-1} + G_{k-1,i} \eta_{k-1,k-1} + \sum_{j=k}^{i-1} \bar{G}_{j,i} \eta_{k-1,j}||_{S^{-1}}^2 \leq 2\delta_i.$$

Hence, at instant k, the constraint of Problem 2 $D(r_i || \bar{r}_i | \mathcal{I}_k) \leq \delta_i$ can be satisfied with $\Omega_k^N = \{\eta_{k-1,k}, 0, \ldots, \eta_{k-1,N}, 0\}$, which means Algorithm 1 is feasible at instant k. Therefore, by induction, the proof is completed. □

Finally, with the following lemma, we prove that the optimal attack is given in Algorithm 1.

Lemma 2. *[24] Let $g(z, \xi)$ be a function such that, for any z, $\min_{\xi \in \Xi} g(z, \xi)$ exists and Ξ is a set of functions such that for every $\xi \in \Xi$, the expectation $\mathbb{E}_z\{g(z, \xi)\}$ exists. Then, we have $\min_{\xi \in \Xi} \mathbb{E}_z\{g(z, \xi)\} = \mathbb{E}_z\{\min_{\xi \in \Xi} g(z, \xi)\}$.*

Theorem 2. *Algorithm 1 provides the optimal attack that solves Problem 1.*

Proof. We resort to dynamic programming to solve Problem 1. According to (11) and (12), the optimal cost function is given as

$$\begin{aligned} J_N^* =& \min_{\eta_N, \Gamma_N} \ \mathbb{E}_{z_{N+1}}\{||F_1 z_{N+1}||_{Q_{N+1}}^2 | \mathcal{I}_N\}, \\ & \text{s.t.} \quad D(r_{N+1} || \bar{r}_{N+1} | \mathcal{I}_N) \leq \delta_{N+1}, \\ J_k^* =& \min_{\eta_k, \Gamma_k} \ \mathbb{E}_{z_{k+1}}\{(||F_1 z_{k+1}||_{Q_{k+1}}^2 + J_{k+1}^*) | \mathcal{I}_k\}, \\ & \text{s.t.} \quad D(r_{k+1} || \bar{r}_{k+1} | \mathcal{I}_k) \leq \delta_{k+1}, k = 0, \ldots, N-1. \end{aligned}$$

Hence, we only need to prove

$$\begin{aligned} J_k^* =& \min_{\Omega_k^N} \ \tilde{J}_k, \\ & \text{s.t.} \ D(r_i || \bar{r}_i | \mathcal{I}_k) \leq \delta_i, i = k+1, \ldots, N+1, \end{aligned} \quad (18)$$

for $k = 0$, …, N.

It is easy to show that (18) holds when $k = N$. Then, we suppose (18) holds for instant $k+1$ and prove it also holds for instant k.

The right-hand-side of (18) equals to

$$\begin{aligned} &\min_{\eta_k, \Gamma_k} \min_{\Omega_{k+1}^N} \ \mathbb{E}_{z_{k+1}} \mathbb{E}_{z_{k+2}^{N+1}} \left\{ \left\{ \sum_{i=k+1}^{N+1} ||F_1 z_i||_{Q_i}^2 | \mathcal{I}_k \right\} \right\}, \\ & \text{s.t.} \quad D(r_i || \bar{r}_i | \mathcal{I}_k) \leq \delta_i, i = k+1, \ldots, N+1. \end{aligned}$$

We use Lemma 2 to exchange $\min_{\Omega_{k+1}^N}$ and $\mathbb{E}_{z_{k+1}}$. Note that the minimization over Ω_{k+1}^N is a function of z_{k+1}. Therefore, the expectation over z_{k+2}^{N+1} is the conditional expectation given $\mathcal{I}_{k+1}$. Then, the-right-hand side of (18) becomes

$$\begin{aligned} &\min_{\eta_k, \Gamma_k} \ \mathbb{E}_{z_{k+1}}\{(||F_1 z_{k+1}||_{Q_{k+1}}^2 + \hat{J}_{k+1}) | \mathcal{I}_k\}, \\ & \text{s.t.} \ D(r_{k+1} || \bar{r}_{k+1} | \mathcal{I}_k) \leq \delta_{k+1}, \end{aligned}$$

where

$$\begin{aligned} \hat{J}_{k+1} =& \min_{\Omega_{k+1}^N} \ \tilde{J}_{k+1}, \\ & \text{s.t.} \ D(r_i || \bar{r}_i | \mathcal{I}_{k+1}) \leq \delta_i, i = k+2, \ldots, N+1. \end{aligned}$$

By the assumption of $\hat{J}_{k+1} = J_{k+1}^*$, we have (18) is true for instant k. Hence, by induction, the proof is completed. □

Remark 3. *At instant k, Problem 2 has $(p+q)(p+q+3)(N-k+1)/2$ scalar decision variables. Moreover, Algorithm 1 requires* *solving Problem 2 repeatedly. Hence, when N is large, Algorithm 1 is computationally expensive. To reduce the computational burden, one can obtain a suboptimal attack by solving the following problem with fewer variables instead of Problem 2 in Algorithm 1 at first few instants:*

$$\begin{aligned}\min_{\Omega_k^{k+W}} \quad & \mathbb{E}_{z_{k+1}^{N+1}}\left\{\sum_{i=k+1}^{N+1}||F_1 z_i||_{Q_i}^2|\mathcal{I}_k\right\},\\ \text{s.t.} \quad & D(r_i||\bar{r}_i|\mathcal{I}_k)\le\delta_i, i=k+1,\ldots,N+1,\\ & \xi_{k+1}^N=0,\end{aligned}$$

where $W\in[0,N-1]$ is a time window.

B. Optimal Strictly Stealthy Attack

In this subsection, we further analyze the case of $\delta_i=0$, $i=1$, $\ldots$, $N+1$, and give the analytical expression of the optimal strictly stealthy attack, which requires much less computational cost than Algorithm 1. By (15) and (16), one of the optimal covariances is 0. Hence, to simplify the problem, we assume the attack is deterministic (i.e., $\xi_k=\eta_k$, $\Gamma_k=0$) and calculate η_k.

Before designing the optimal strictly stealthy attack, we first give the lemma which characterizes the feasible set of the strictly stealthy attack.

Lemma 3. *When $\delta_i=0$, $i=1$, $\ldots$, $N+1$, the attack should be in the following set:*

$$\eta_k\in\{\eta|\bar{\Upsilon}_k F_2 z_k+G_k\eta=0\}, k=0,\ldots,N, \tag{19}$$

where

$$\begin{aligned}\bar{\Upsilon}_N &= CA, G_N=G,\\ \bar{\Upsilon}_k &= \begin{bmatrix}CA\\ \Upsilon_{k+1}\bar{A}\end{bmatrix}, G_k=\begin{bmatrix}G\\ \Upsilon_{k+1}\bar{B}\end{bmatrix}, k=0,\ldots,N-1,\\ \Upsilon_k &= \Psi_k^T\Psi_k-\Psi_k^T\Phi_k(\Phi_k^T\Phi_k)^\dagger\Phi_k^T\Psi_k,\\ \Psi_k &= \begin{bmatrix}(CA)^T & (CA\bar{A})^T & \cdots & (CA\bar{A}^{N-k})^T\end{bmatrix}^T,\\ k &= 0,\ldots,N,\\ \Phi_N &= G,\\ \Phi_k &= \begin{bmatrix}G & 0 & \cdots & 0\\ CA\bar{B} & G & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ CA\bar{A}^{N-k-1}\bar{B} & CA\bar{A}^{N-k-2}\bar{B} & \cdots & G\end{bmatrix},\\ k &= 0,\ldots,N-1.\end{aligned}$$

Proof. From (9), to be strictly stealthy, η_k should satisfy

$$CA\Delta e_k+G\eta_k=0. \tag{20}$$

When $k=N$, we can directly obtain (19) from (20). When $k<N$, η_k should not only satisfy (20), but also be chosen such that there exists a strictly stealthy attack from instants $k+1$ to N, i.e., there exists an η_{k+1}^N such that $CA\Delta e_i+G\eta_i=0$, $i=k+1$, $\ldots$, N. Then, from (10), there exists an η_{k+1}^N such that

$$\Psi_{k+1}\Delta e_{k+1}+\Phi_{k+1}\eta_{k+1}^N=0, \tag{21}$$

which is equivalent to

$$\begin{aligned}\min_{\eta_{k+1}^N}||\Psi_{k+1}\Delta e_{k+1}+\Phi_{k+1}\eta_{k+1}^N||_{I_n}^2 &= ||\Delta e_{k+1}||_{\Upsilon_{k+1}}^2\\ &= 0. \end{aligned}\tag{22}$$

Combining (10), (20) and (22), we can obtain (19) for $k=0$, $\ldots$, $N-1$, which completes the proof. □

It can be seen from Lemma 3 that (19) is satisfied when $\eta_0^N=0$. However, $\eta_0^N=0$ does not have any influence on the system. Hence, we give the necessary and sufficient condition for the existence of nonzero η_0^N in the following corollary.

Corollary 1. *A nonzero η_0^N exists if and only if G does not have full column rank.*

Proof. Recalling that $\Delta e_0=0$, it follows from (21) (by letting $k=-1$) that the necessary and sufficient condition for the existence of nonzero η_0^N is that Φ_0 does not have full column rank, which holds if and only if G does not have full column rank (since Φ_0 is a block lower triangular matrix with G in the diagonal). □

Since $G\in\mathbf{R}^{m\times(p+q)}$, we can also obtain the following sufficient condition:

Corollary 2. *If $p+q>m$, i.e., the dimension of ξ_k is larger than that of y_k, a nonzero η_0^N exists.*

We then give the lemma which will be used to derive the analytical expression of the optimal strictly stealthy attack.

Lemma 4. *Suppose γ is a known vector, then, for $k=0$, $\ldots$, N, the following equation*

$$\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k\alpha+\mathcal{G}_k^T\tilde{B}^T\gamma=0 \tag{23}$$

with

$$\begin{aligned}\mathcal{G}_k &= I_{p+q}-G_k^\dagger G_k, k=0,\ldots,N,\\ \mathcal{Q}_{N+1} &= F_1^T Q_{N+1}F_1,\\ \mathcal{Q}_k &= F_1^T Q_k F_1+(\bar{A}-G_k^\dagger\bar{\Upsilon}_k F_2)^T\mathcal{Q}_{k+1}(\bar{A}-G_k^\dagger\bar{\Upsilon}_k F_2)\\ &\quad -\Xi_k^T(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)^\dagger\Xi_k,\\ \Xi_k &= \mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}(\tilde{A}-G_k^\dagger\bar{\Upsilon}_k F_2),\\ k &= 1,\ldots,N,\end{aligned}$$

has at least one solution α.

Proof. We prove the solution exists by proving $\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)=\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T)$. Trivially, we have $\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)\subset\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T)$. Hence, we only need to show $\mathrm{Null}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)\subset\mathrm{Null}(\tilde{B}\mathcal{G}_k)$ by contradiction.

Suppose $\alpha\in\mathrm{Null}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)$ while $\alpha\notin\mathrm{Null}(\tilde{B}\mathcal{G}_k)$. Then, we have $\mathcal{G}_k\alpha=\begin{bmatrix}\alpha_1^T & \alpha_2^T\end{bmatrix}^T\neq0$, where $\alpha_1\in\mathbf{R}^p$. Furthermore, we have

$$\begin{aligned}G_N\mathcal{G}_N\alpha &= CD\alpha_1+E\alpha_2,\\ G_k\mathcal{G}_k\alpha &= \begin{bmatrix}CD\alpha_1+E\alpha_2\\ \Upsilon_{k+1}\bar{B}\mathcal{G}_k\alpha\end{bmatrix}, k=0,\ldots,N-1.\end{aligned}$$

Note that E has full column rank. If $\alpha_1=0$, then $CD\alpha_1+E\alpha_2\neq0$ (since $\begin{bmatrix}\alpha_1^T & \alpha_2^T\end{bmatrix}^T\neq0$), which contradicts the fact that $G_k\mathcal{G}_k=0$. Hence, one has $\alpha_1\neq0$.

Since $Q_{k+1}\succ0$ and D has full column rank, we have

$$\begin{aligned}\alpha^T\mathcal{G}_k^T\tilde{B}^T F_1^T Q_{k+1}F_1\tilde{B}\mathcal{G}_k\alpha &= \alpha_1^T D^T Q_{k+1}D\alpha_1\\ &> 0.\end{aligned}$$

Moreover, it is easy to show that $\mathcal{Q}_{k+1}-F_1^T Q_{k+1}F_1\succeq0$. Then, it follows that

$$\alpha^T\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k\alpha>0, \tag{24}$$

which is contrary to the assumption of $\alpha\in\mathrm{Null}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)$.

Hence, we have $\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)=\mathrm{Span}(\mathcal{G}_k^T\tilde{B}^T)$, which completes the proof. □

We finally give the lemma which will be used to prove that the optimal strictly stealthy attack is unique.

Lemma 5. *Suppose α_3 and α_4 are two solutions of* (23)*, then $\mathcal{G}_k\alpha_3 = \mathcal{G}_k\alpha_4$.*

Proof. If $\mathcal{G}_k\alpha_3 \neq \mathcal{G}_k\alpha_4$, according to (24), we have $||\mathcal{G}_k(\alpha_3 - \alpha_4)||^2_{\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}} > 0$, which contradicts the fact that $\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k\alpha_3 = -\mathcal{G}_k^T\tilde{B}^T\gamma = \mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k\alpha_4$. Hence, the proof is completed. □

Now, we are ready to design the optimal strictly stealthy attack.

Theorem 3. *The optimal attack policy $\mathcal{F}^*(\mathcal{I}_k)$ that solves Problem 1 with $\delta_i = 0$, $i = 1, \ldots, N+1$, is $\xi_k^* = \eta_k^*$, where*

$$\begin{aligned}\eta_k^* &= -G_k^\dagger\bar{\Upsilon}_kF_2z_k - \mathcal{G}_k(\mathcal{G}_k^T\tilde{B}^T\mathcal{Q}_{k+1}\tilde{B}\mathcal{G}_k)^\dagger\Xi_kz_k,\\ k &= 0, \ldots, N,\end{aligned} \tag{25}$$

with G_k, $\bar{\Upsilon}_k$ defined in Lemma 3 and $\mathcal{G}_k$, $\mathcal{Q}_{k+1}$, Ξ_k defined in Lemma 4.

Proof. Similar to the proof of Theorem 2, the optimal cost function is given as:

$$\begin{aligned}J_N^* &= \min_{\eta_N} \mathbb{E}\{||F_1z_{N+1}||^2_{Q_{N+1}}|\mathcal{I}_N\},\\ &\quad \text{s.t. } \bar{\Upsilon}_NF_2z_N + G_N\eta_N = 0,\end{aligned} \tag{26}$$

$$\begin{aligned}J_k^* &= \min_{\eta_k} \mathbb{E}\{(||F_1z_{k+1}||^2_{Q_{k+1}} + J_{k+1}^*)|\mathcal{I}_k\},\\ &\quad \text{s.t. } \bar{\Upsilon}_kF_2z_k + G_k\eta_k = 0, k = 0, \ldots, N-1.\end{aligned} \tag{27}$$

We begin by finding the optimal attack at instant N. By (26), we have $\eta_N^* = -G_N^\dagger\bar{\Upsilon}_NF_2z_N + \mathcal{G}_Nc^*$, where c^* satisfies

$$\mathcal{G}_N^T\tilde{B}^T\mathcal{Q}_{N+1}(\tilde{A}z_N + \tilde{B}(-G_N^\dagger\bar{\Upsilon}_NF_2z_N + \mathcal{G}_Nc^*)) = 0. \tag{28}$$

It follows from Lemma 4 that the solution of (28) exists. It is easy to get one solution

$$c^* = -(\mathcal{G}_N^T\tilde{B}^T\mathcal{Q}_{N+1}\tilde{B}\mathcal{G}_N)^\dagger\Xi_Nz_N.$$

Hence, the optimal strictly stealthy attack (25) with $k = N$ can be obtained. Note that, by Lemma 5, all the solutions of (28) lead to the same $\mathcal{G}_Nc^*$. Therefore, the optimal strictly stealthy attack is unique and

$$\begin{aligned}J_N^* &= z_N^T[(\bar{A} - G_N^\dagger\bar{\Upsilon}_NF_2)^T\mathcal{Q}_{N+1}(\bar{A} - G_N^\dagger\bar{\Upsilon}_NF_2)\\ &\quad -\Xi_N^T(\mathcal{G}_N^T\tilde{B}^T\mathcal{Q}_{N+1}\tilde{B}\mathcal{G}_N)^\dagger\Xi_N]z_N + \text{Tr}(\mathcal{Q}_{N+1}H^TSH).\end{aligned}$$

Furthermore, in (27), let $k = N-1$ and substitute J_N^* into (27). Then, we have

$$\begin{aligned}J_{N-1}^* = \min_{\eta_{N-1}} \mathbb{E}\{(||z_N||^2_{\mathcal{Q}_N} + \text{Tr}(\mathcal{Q}_{N+1}H^TSH))|\mathcal{I}_{N-1}\},\\ \text{s.t. } \bar{\Upsilon}_{N-1}F_2z_{N-1} + G_{N-1}\eta_{N-1} = 0.\end{aligned}$$

Note that $\text{Tr}(\mathcal{Q}_{N+1}H^TSH)$ is a constant. Following the same method as that used to find η_N^*, we can get η_{N-1}^*.

Repeating the dynamic programming procedure for $k = N-2$, $\ldots$, 0, we can obtain that the optimal attack has the form of (25) and is unique. The proof is completed. □

Remark 4. *It can be seen that the optimal strictly stealthy attack is a state feedback of z_k, which can be calculated online by the attacker based on the information $\mathcal{I}_k$ (see Remark 2). Moreover, it is easy to find from* (25) *that when G_k has full column rank for all k, then $\eta_k^* = 0$. This coincides with the discussion in Corollary 1, since $G_N = G$ and $G_k = \begin{bmatrix} G \\ \Upsilon_{k+1}\bar{B}\end{bmatrix}$, $k = 0, \ldots, N-1$.*

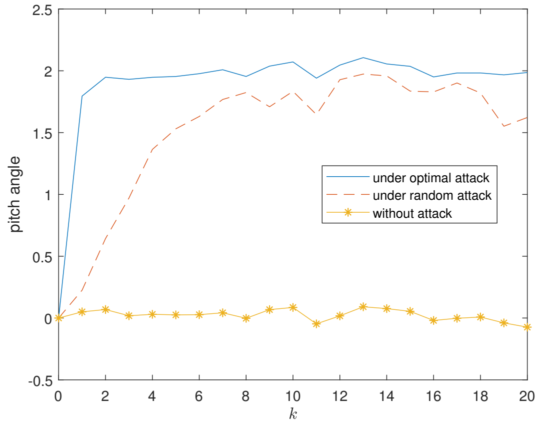

Fig. 2. The pitch angles under optimal strictly stealthy attack, a random generated strictly stealthy attack and no attack

IV. Numerical Example

In this section, the example of a simplified longitudinal flight control system [30] is given to illustrate the effectiveness of the proposed attack.

We use the linearized discrete-time model, where the state x_k represents the pitch angle, the pitch rate and the velocity of the flight. The parameters of system (1) and (2) are given by

$$A = \begin{bmatrix} 0.9944 & -0.1203 & -0.4302 \\ 0.0017 & 0.9902 & -0.0747 \\ 0 & 0.8187 & 0 \end{bmatrix}, B = \begin{bmatrix} 0.4252 \\ -0.0082 \\ 0.1813 \end{bmatrix},$$

$C = E = I_3$, $D = B$, $\Sigma_w = \Sigma_v = 0.001I_3$.

The controller gain is chosen to be

$$L = \begin{bmatrix} 0.6311 & -0.0136 & -0.0239 \\ -0.0136 & 0.5846 & 0.1287 \\ -0.0239 & 0.1287 & 0.2876 \end{bmatrix}.$$

Moreover, the initial state is assumed to be $x_0 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$.

The attacker's goal is to move the state of the flight to $x^* = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}^T$. The time horizon in Problem 1 is $N = 19$ and the weight matrices are $Q_i = I_3$, $i = 1, \ldots, 20$.

a) Strictly stealthy attack : Let $E = I_3$, meaning that all sensors can be corrupted. According to Corollary 1, a nonzero strictly stealthy attack exists since $G = \begin{bmatrix} CD & E \end{bmatrix}$ does not have full column rank. The pitch angles of the flight under optimal strictly stealthy attack (25), under a randomly generated strictly stealthy attack $\xi_k' = -G_k^\dagger CAF_2z_k + \mathcal{G}_kc_k$ with $c_k \sim \mathcal{N}(0, 0.1I_4)$ and under no attack are shown in Fig.2, which indicates that under the attack (25), the system's state can reach x^*.

b) Non-strictly stealthy attack : When E equals to $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^T$ (i.e., the last sensor cannot be corrupted), the nonzero strictly stealthy attack does not exist. We choose the thresholds in Problem 1 to be $\delta_k = 10$, $k = 1, \ldots, 20$. The pitch angles of the flight and the values of $D(r_k||\bar{r}_k)$ under optimal stealthy attack calculated by Algorithm 1, under a randomly generated attack $\xi_k' \sim \mathcal{N}(\epsilon_k\eta_{k,k}, \epsilon_k^2\Gamma_{k,k})$ with $\epsilon_k \in (0, 1)$ and under no attack are given in Fig.3 and Fig.4, respectively. Fig.3 and Fig.4 show that the attack given in Algorithm 1 is stealthy and can effectively make the system reach the target state. Fig. 5 presents the value of $\sum_{i=1}^{N+1}||x_i - x^*||^2_{Q_i}$ with different δ_k. It is observed that with larger δ_k, the average distance between the state and the target is smaller.

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2022.3192201

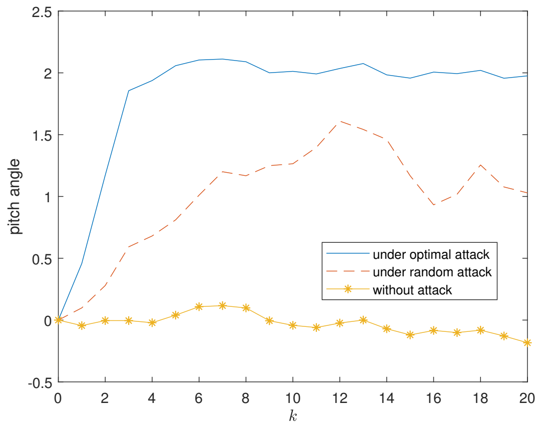

Fig. 3. The pitch angles under optimal stealthy attack with $\delta_k = 10$, a random generated attack and no attack

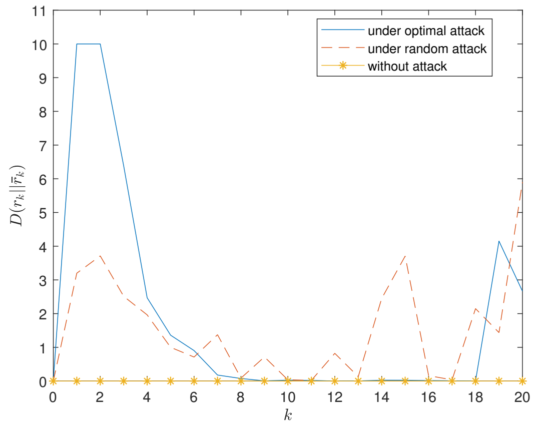

Fig. 4. The values of $D(r_k||\bar{r}_k)$ under optimal stealthy attack with $\delta_k = 10$, a random generated attack and no attack

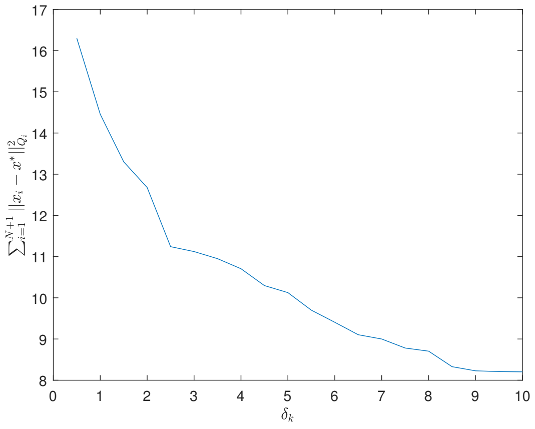

Fig. 5. The values of $\sum_{i=1}^{N+1} ||x_i - x^*||^2_{Q_i}$ with different δ_k

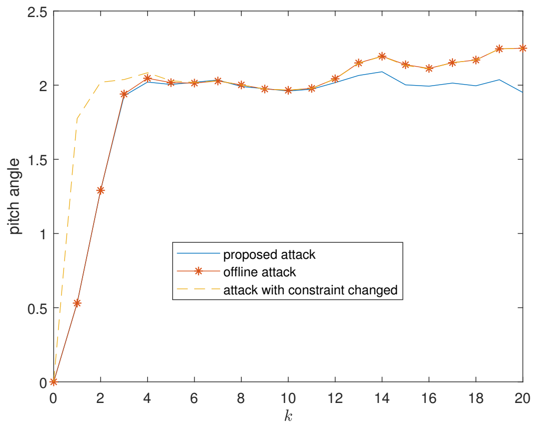

Fig. 6. Comparison among different attack strategies

Finally, we compare Algorithm 1 with two attack strategies in Fig. 6. The first one is an offline attack, i.e., the attack ξ_0^N obtained by solving Problem 2 for $k = 0$. It is shown that the offline attack and the attack computed by Algorithm 1 make the state reach the target almost simultaneously. However, after instant 12, the pitch angle under the offline attack deviates more far from 2 due to the influence of the noise. The second one is the optimal stealthy attack with constraint in Problem 1 changed by $D(r_1^{N+1}||\bar{r}_1^{N+1}) \le \sum_{k=1}^{N+1} \delta_k$. Although this attack is also computed offline (as discussed in Remark 1), it drives the state to the target more quickly than our proposed attack, which is because the constraint used in Problem 1 restricts the stealthiness of the attack for each instant, not a period of time.

V. Conclusion

In this article, we have studied the design of online KLD-based stealthy attack against CPSs for two cases. We have shown that the optimal attack is a solution of a convex optimization problem that should be solved by the attacker at each instant. The feasibility of the attack policy has also been discussed. When the threshold equals to zero, we have proved there always exists a unique optimal strictly stealthy attack, which is a state feedback of an augmented system run by the attacker. An example of a simplified longitudinal flight control system has been presented to show the effect of the attack. Future works may involve expanding the results to large-scale systems.

References

[1] P. Derler, E. A. Lee, and A. S. Vincentelli, “Modeling cyber–physical systems,” *Proceedings of the IEEE*, vol. 100, no. 1, pp. 13–28, 2012.

[2] K.-D. Kim and P. R. Kumar, “Cyber–physical systems: A perspective at the centennial,” *Proceedings of the IEEE*, vol. 100, pp. 1287–1308, 2012.

[3] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “A secure control framework for resource-limited adversaries,” *Automatica*, vol. 51, pp. 135–148, 2015.

[4] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, “A systems and control perspective of CPS security,” *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

[5] A. Humayed, J. Lin, F. Li, and B. Luo, “Cyber-physical systems security–A survey,” *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802–1831, 2017.

[6] M. Wolf and D. Serpanos, “Safety and security in cyber-physical systems and internet-of-things systems,” *Proceedings of the IEEE*, vol. 106, no. 1, pp. 9–20, 2018.

[7] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, “A survey of physics-based attack detection in cyber-physical systems,” *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, p. 76, 2018.

[8] P. Griffioen, S. Weerakkody, B. Sinopoli, O. Ozel, and Y. Mo, "A tutorial on detecting security attacks on cyber-physical systems," in *Proceedings of the 18th European Control Conference (ECC)*, pp. 979–984, 2019.

[9] Y. Chen, S. Kar, and J. M. Moura, "Cyber-physical attacks with control objectives," *IEEE Transactions on Automatic Control*, vol. 63, no. 5, pp. 1418–1425, 2017.

[10] C. Murguia and J. Ruths, "On model-based detectors for linear time-invariant stochastic systems under sensor attacks," *IET Control Theory & Applications*, vol. 13, no. 8, pp. 1051–1061, 2019.

[11] D. Umsonst and H. Sandberg, "Anomaly detector metrics for sensor data attacks in control systems," in *Proceedings of the 2018 Annual American Control Conference (ACC)*, pp. 153–158, 2018.

[12] C. Kwon and I. Hwang, "Reachability analysis for safety assurance of cyber-physical systems against cyber attacks," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2272–2279, 2017.

[13] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[14] Y. Mao, H. Jafarnejadsani, P. Zhao, E. Akyol, and N. Hovakimyan, "Novel stealthy attack and defense strategies for networked control systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3847–3862, 2020.

[15] T. Sui, Y. Mo, D. Marelli, X. Sun, and M. Fu, "The vulnerability of cyber-physical system under stealthy attacks," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 637–650, 2021.

[16] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[17] J. Milošević, A. Teixeira, K. H. Johansson, and H. Sandberg, "Actuator security indices based on perfect undetectability: Computation, robustness, and sensor placement," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3816–3831, 2020.

[18] C.-Z. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, 2017.

[19] Q. Zhang, K. Liu, D. Han, G. Su, and Y. Xia, "Design of stealthy deception attacks with partial system knowledge," *IEEE Transactions on Automatic Control*, DOI:10.1109/TAC.2022.3146079.

[20] Q. Zhang, K. Liu, Y. Xia, and A. Ma, "Optimal stealthy deception attack against cyber-physical systems," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3963–3972, 2020.

[21] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117–124, 2018.

[22] X.-X. Ren and G.-H. Yang, "Kullback-Leibler divergence-based optimal stealthy sensor attack against networked linear quadratic Gaussian systems," *IEEE Transactions on Cybernetics*, DOI: 10.1109/TCYB.2021.3068220.

[23] Q. Zhang, K. Liu, Z. Pang, Y. Xia, and T. Liu, "Reachability analysis of cyber-physical systems under stealthy attacks," *IEEE Transactions on Cybernetics*, DOI: 10.1109/TCYB.2020.3025307.

[24] J. L. Speyer and W. H. Chung, *Stochastic processes, estimation, and control*. SIAM, 2008.

[25] Z. Guo, D. Shi, D. E. Quevedo, and L. Shi, "Secure state estimation against integrity attacks: A Gaussian mixture model spproach," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 194–207, 2018.

[26] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[27] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[28] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*. SIAM, 1994.

[29] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[30] L. Hu, Z. Wang, Q. L. Han, and X. Liu, "State estimation under false data injection attacks: Security analysis and system protection," *Automatica*, vol. 87, pp. 176–183, 2018.