Reachability Analysis of Cyber-Physical Systems Under Stealthy Attacks

Qirui Zhang, Kun Liu, *Member, IEEE*, Zhonghua Pang, *Senior Member, IEEE*, Yuanqing Xia, *Senior Member, IEEE*, and Tao Liu

***Abstract*—This article studies the reachable set of cyber-physical systems subject to stealthy attacks with the Kullback–Leibler divergence adopted to describe the stealthiness. The reachable set is defined as the set in which both the system state and the estimation error of the Kalman filter reside with a certain probability. The necessary and sufficient conditions of the reachable set being unbounded are given for the finite and infinite time cases, respectively. When the reachable set is bounded, an ellipsoidal outer approximation is obtained by solving a convex optimization problem. An application of this approximation to the safety evaluation is also given. A numerical simulation of an unmanned ground vehicle is presented to demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Cyber-physical system (CPS) security, Kullback–Leibler divergence (KLD), reachable set, stealthy attack.**

I. Introduction

With the rapid development of communication, computation, and control technology, cyber-physical systems (CPSs) have become a hot research topic in both academia and applications [1], [2]. Due to the openness of the network, CPSs are more vulnerable to malicious threats, which may result in economic losses and physical damages. Therefore, it is of vital importance to investigate the issue of CPS security, which has received considerable attention in recent years [3]–[6].

Research on CPS security mainly involves methods of detecting attack [7]–[9]; games between the attacker and the defender [10], [11]; protection strategies against attacks [12], [13]; and performance degradations resulting from attacks, including control performance [14], [15]; estimation error [16]; reachable set [17]–[22]; etc.

The CPS is usually equipped with attack detectors, while some carefully designed attacks can avoid being detected. For example, in 2010, using the Stuxnet worm, the attacker targeting a uranium enriching facility in Iran, injected stealthy data to vary rotational speeds, and made the rotors of centrifuges break [23]. Studying the reachable set of the CPS subject to stealthy attacks helps us to redesign system parameters so that the CPS can avoid being driven to critical states. For the χ^2 detector, the necessary and sufficient condition for the reachable set of the CPS suffering stealthy sensor attacks being unbounded is given in [17], and the case where both actuator attacks and sensor attacks exist is further discussed in [18]. In [19], for a class of the residual-based detector, an enforcement policy is provided to keep the reachable set bounded against sensor attacks. Moreover, to approximate the reachable set, an ellipsoidal inner and outer approximation technique is developed in [20] for the system with the χ^2 detector. An algorithm that computes the exact reachable set is proposed in [21], where the detector is based on the sequential probability ratio test. For a class of energy-based detectors, a linear matrix inequality (LMI) approach is provided in [22] to approximate the reachable set with an outer ellipsoid. Nevertheless, these works only consider stealthy attacks against specific detectors.

The Kullback–Leibler divergence (KLD) is usually used to describe the stealthiness of the attacks with respect to arbitrary detectors. The relationship between the KLD and the false alarm rate (the rate that the detector alarms when there is no attack) is discussed in [24]. The work of [24] is extended to the rate of successfully detecting attacks in [25]. With the KLD adopted as the stealthy constraint of the attack, the control performance and the estimation performance degradations are investigated in [26], [27] and [16], [28], respectively. The zero-sum game between the attacker and the defender with the false alarm rate as the reward is studied in [29]. It is proved that the equilibrium attack strategy is the one that minimizes the KLD between legitimate and falsified states. However, the reachable set with the KLD used to describe stealthiness is yet to be investigated, and the methods to study the reachable set in the above literature are not applicable.

Therefore, in this article, with the KLD as the metric of stealthiness, we analyze the reachability of the system with a Kalman filter and a state-feedback controller. To remain stealthy, the attacker should keep the KLD between the filter innovations without and with attack no smaller than a

Manuscript received July 9, 2020; accepted September 13, 2020. Date of publication October 29, 2020; date of current version June 16, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61873034 and Grant 61673023; in part by the Natural Science Foundation of Beijing Municipality under Grant 4182057; in part by the Open Subject of Beijing Intelligent Logistics System Collaborative Innovation Center under Grant BILSCIC-2019KF-13; and in part by the Graduate Technological Innovation Project of Beijing Institute of Technology under Grant 2019CX20031. This article was recommended by Associate Editor M. Wu. *(Corresponding author: Kun Liu.)*

Qirui Zhang, Kun Liu, and Yuanqing Xia are with the School of Automation, Beijing Institute of Technology, Beijing 100081, China (e-mail: qiruizhang@bit.edu.cn; kunliubit@bit.edu.cn; xia_yuanqing@bit.edu.cn).

Zhonghua Pang is with the School of Electrical and Control Engineering, North China University of Technology, Beijing 100043, China (e-mail: zhonghua.pang@ia.ac.cn).

Tao Liu is with the School of Information, Beijing Wuzi University, Beijing 101149, China (e-mail: liutao19832001@163.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.3025307.

Digital Object Identifier 10.1109/TCYB.2020.3025307

2168-2267 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

threshold. The reachable set is defined as the set in which both the system state and the estimation error of the Kalman filter reside with a certain probability. The contributions of this article are summarized as follows.

1) The necessary and sufficient conditions of the reachable set being unbounded are given for the finite and infinite time cases, respectively.
2) The ellipsoidal outer approximation for the bounded reachable set is computed by solving a convex optimization problem.
3) The safety of the system is evaluated as an application of this approximation.

The remainder of this article is organized as follows. Section II gives a description of the problem we consider. In Section III, the boundedness of the reachable set is discussed. In Section IV, a convex optimization problem is solved to compute the outer ellipsoid that approximates the reachable set and the safety of the system is evaluated. Section V presents a simulation to illustrate the effectiveness of our proposed approximation approach. Section VI 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. $\mathbf{N}$ stands for the set of natural numbers and $\mathbf{R}_{\geq 0}$ is the set of non-negative real numbers. $x \sim \mathcal{N}(a, \Sigma)$ means the vector x satisfies a Gaussian distribution with mean value a and covariance matrix Σ. For any matrix A, $A^\dagger$ is the Moore–Penrose pseudoinverse, Rank(A) represents the rank of A, Null(A) stands for the null space of A, the column space of A is denoted by Span(A), and Tr(A) is used to denote the trace of A when A is a square matrix. $\|\cdot\|$ stands for the Euclidean norm. blkdiag(A, B) is the block-diagonal matrix with matrices A and B in the diagonal. For any symmetric matrix P, the notation $P \succ 0$ ($P \succeq 0$) means that P is positive definite (semidefinite).

II. Problem Formulation

In this section, we present the problem setup that we focus on.

A. System Model

Consider the following discrete-time system:

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

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

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. A, B, C, D, and E are real-valued matrices of compatible dimensions. 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. Both D and E are injective.

The system runs without knowing the attack signals u_k^a and y_k^a. A Kalman filter is used to estimate the state. It is well known that the Kalman filter converges exponentially fast from any initial condition. Hence, without loss of generality, we assume that the filter starts from the steady state, which makes the filter gain fixed, that is, the Kalman filter has the form

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

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

where $\hat{x}_k$ is the estimate of x_k, $K = PC^T(CPC^T + \Sigma_v)^{-1}$, and $P = APA^T + \Sigma_w - APC^T(CPC^T + \Sigma_v)^{-1}CPA^T$. It is assumed that $A - KCA$ is stable.

We 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. Let $\bar{r}_k$ be the innovation of the Kalman filter when the system is under no attack, then $\bar{r}_k$ has i.i.d. $\mathcal{N}(0, Q)$ with $Q = CPC^T + \Sigma_v$.

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

The KLD, which 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) [30]: 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

$$D(x||y) = \int_{\{\zeta | f_x(\zeta) > 0\}} f_x(\zeta) \log \frac{f_x(\zeta)}{f_y(\zeta)} d\zeta. \tag{6}$$

In this article, we do not consider the attack being stealthy against any specific detectors. Instead, following [24]–[29], we use the KLD between the filter innovations without and with attack $D(\bar{r}_k||r_k)$ to describe the stealthiness of the attack. To keep stealthy, the attacker should let $D(\bar{r}_k||r_k)$ be no larger than a positive threshold δ.

Without loss of generality, the actuator attack and the sensor attack are assumed to start at time instants 0 and 1, respectively. For the attacker, we also make the following assumptions.

Assumption 1 (Attacker's Knowledge): The attacker has full knowledge of the CPS, that is, the attacker knows the matrices A, B, C, D, E, L, Σ_w, and Σ_v.

Assumption 2 (Attack Signal): The attacker injects attack signals u_k^a and y_{k+1}^a such that the innovation $r_k \sim \mathcal{N}(\eta_k, \Sigma_k^{-1})$ with mean value $\eta_k \in \mathbf{R}^m$ and covariance $\Sigma_k^{-1} \succ 0$.

Remark 1: Assumption 1 is commonly used in the existing literature [14]–[22], [24]–[28]. It is necessary since it allows the attacker to carefully design attacks and to keep itself stealthy. Moreover, it is reasonable since the attacker may acquire the parameters of the CPS from specific physical problems or by system identification techniques [31]. Furthermore, Assumption 2 includes several kinds of attack scenarios, e.g., the deterministic attack signals in [14] and [19], the Gaussian attack signals in [24] and [27], and the innovation-based attack in [16] and [28]. It should be pointed out that the attack signals may be dependent of the noises. Hence, $\Sigma_k^{-1} \succeq Q$ is not necessarily true.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

C. Definition of the Reachable Set

Let the integer $N > 0$ be the time horizon and $z_k = \begin{bmatrix} x_k^T & e_k^T \end{bmatrix}^T$. In this article, we are interested in the reachable set of z_k at time instant N, that is, the set containing all the possible values of z_N under stealthy attack. Then, we can define the reachable set by $\text{Reach}(z_N, \psi) = \{z_N | z_0 = \psi, D(\bar{r}_k || r_k) \le \delta, k = 1, \ldots, N\}$, where $\psi = \begin{bmatrix} \psi_x^T & \psi_e^T \end{bmatrix}^T$ with ψ_x and ψ_e being the initial state and estimation error, respectively.

Note that the noises w_k and v_k are Gaussian, which means that z_N can be any vector. Hence, instead of $\text{Reach}(z_N, \psi)$, we consider another reachable set defined according to the statistical characteristics of the system.

Definition 2: The reachable set of z_k at time instant N is

$$\begin{aligned} Rch(z_N, \psi) = \big\{ z_N | z_0 = \psi, (r_k - \eta_k)^T \Sigma_k (r_k - \eta_k) \le a \\ w_{k-1}^T \Sigma_w^{-1} w_{k-1} \le b, v_k^T \Sigma_v^{-1} v_k \le c \\ D(\bar{r}_k || r_k) \le \delta, k = 1, \ldots, N \big\} \end{aligned} \tag{7}$$

where a, b, and c are positive constants.

Compared with $\text{Reach}(z_N, \psi)$, Definition 2 includes three additional constraints, that is, the innovation, the process noise, and the measurement noise are in ellipsoids $(r_k - \eta_k)^T \Sigma_k (r_k - \eta_k) \le a$, $w_{k-1}^T \Sigma_w^{-1} w_{k-1} \le b$, and $v_k^T \Sigma_v^{-1} v_k \le c$, respectively. Notice that $(r_k - \eta_k)^T \Sigma_k (r_k - \eta_k)$, $w_{k-1}^T \Sigma_w^{-1} w_{k-1}$, and $v_k^T \Sigma_v^{-1} v_k$ satisfy the χ^2 distribution with m, n, and m degrees of freedom, respectively. Then, the probabilities that r_k, w_{k-1}, and v_k are in corresponding ellipsoids are $1 - F(a, m)$, $1 - F(b, n)$, and $1 - F(c, m)$, respectively, where F is the cumulative distribution function of the χ^2 distribution. Hence, the reachable set (7) is defined as the set in which z_N resides with a certain probability. Moreover, since both m and n are fixed, the parameters a, b, and c can be determined according to the desired probabilities $1 - F(a, m)$, $1 - F(b, n)$, and $1 - F(c, m)$, respectively.

In the rest of this article, we study the boundedness of $Rch(z_N, \psi)$ (i.e., whether $\|z_N\|$ can be $+\infty$) and approximate $Rch(z_N, \psi)$ when it is bounded.

III. Unbounded Reachable Set

In this section, we analyze the boundedness of $Rch(z_N, \psi)$ for finite N and $N \to +\infty$, respectively.

From (1)–(5), it follows that:

$$z_{k+1} = \tilde{A} z_k + \tilde{B} \xi_k + G w_k + H v_{k+1} \tag{8}$$

$$r_{k+1} = \tilde{C} z_k + \tilde{D} \xi_k + C w_k + v_{k+1} \tag{9}$$

where

$$\xi_k = \begin{bmatrix} u_k^a \\ y_{k+1}^a \end{bmatrix}, \quad G = \begin{bmatrix} I_n \\ I_n - KC \end{bmatrix}, \quad H = \begin{bmatrix} 0 \\ -K \end{bmatrix}$$

$$\tilde{A} = \begin{bmatrix} A + BL & -BL \\ 0 & A - KCA \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} D & 0 \\ D - KCD & -KE \end{bmatrix}$$

$$\tilde{C} = \begin{bmatrix} 0 & CA \end{bmatrix}$$

and $\tilde{D} = \begin{bmatrix} CD & E \end{bmatrix}$.

Since (8) and (9) are linear, the impacts of the noises and the attacks on the system can be separated. We use Δe_k and Δr_k to denote the estimation error and the innovation induced by the attacks, respectively, that is, Δe_k and Δr_k are vectors that satisfy

$$\Delta e_{k+1} = (A - KCA) \Delta e_k + \bar{B} \xi_k \tag{10}$$

$$\Delta r_{k+1} = CA \Delta e_k + \tilde{D} \xi_k \tag{11}$$

where $\Delta e_0 = 0$ and $\bar{B} = \begin{bmatrix} D - KCD & -KE \end{bmatrix}$.

Moreover, one has

$$\begin{aligned} e_{k+1} - \Delta e_{k+1} = (A - KCA)(e_k - \Delta e_k) + (I_n - KC) w_k \\ - K v_{k+1} \end{aligned} \tag{12}$$

$$\bar{r}_{k+1} = CA(e_k - \Delta e_k) + C w_k + v_{k+1} \tag{13}$$

where $e_k - \Delta e_k$ and $\bar{r}_k$ are the estimation error and the innovation resulting from noises (i.e., the estimation error and the innovation when there is no attack), respectively.

Then, the following lemma can be obtained.

Lemma 1: Under Assumptions 1 and 2, $Rch(z_N, \psi)$ is unbounded if and only if the set

$$Rch(\Delta e_N, 0) = \{\Delta e_N | \Delta e_0 = 0, \|\Delta r_k\| \le 1, k = 1, \ldots, N\}$$

is unbounded.

Proof: Suppose the eigenvalues of the matrix $\Sigma_k Q$ are κ_i, $i = 1, \ldots, m$. Recall that both r_k and $\bar{r}_k$ are Gaussian distributed. From Definition 1 and $D(\bar{r}_k || r_k) \le \delta$, it follows that:

$$\begin{aligned} D(\bar{r}_k || r_k) &= \frac{1}{2} \big[\text{Tr}(\Sigma_k Q) - \log(|Q||\Sigma_k|) - m + \eta_k^T \Sigma_k \eta_k \big] \\ &= \frac{1}{2} \left[\sum_{i=1}^{m} (\kappa_i - \log(\kappa_i)) - m + \eta_k^T \Sigma_k \eta_k \right] \\ &\le \delta. \end{aligned} \tag{14}$$

Note that $\kappa_i > 0$, $i = 1, \ldots, m$, since $\Sigma_k Q \succ 0$. It can be observed that $\kappa_i - \log(\kappa_i) \ge 1$, $i = 1, \ldots, m$. Hence, $0 \le \eta_k^T \Sigma_k \eta_k \le 2\delta$, which means that $\|\eta_k\|$ is bounded. Furthermore, $\|r_k - \eta_k\|$, $\|w_{k-1}\|$, and $\|v_k\|$ are bounded according to Definition 2. Then, from (12) and (13), we know that $\|e_k - \Delta e_k\|$ and $\|\bar{r}_k\|$ are bounded from time instant 1 to $+\infty$ for the reason that the matrix $A - KCA$ is stable.

Notice that $\Delta r_k = \eta_k + r_k - \eta_k - \bar{r}_k$. Therefore, $\|\Delta r_k\|$ is bounded. Without loss of generality, we set the bound to be 1, that is, $\|\Delta r_k\| \le 1$. By (3)–(5), $\|\hat{x}_k\|$ is always bounded due to the stable matrix $A + BL$ and bounded $\|r_k\|$. Note that

$$z_k = \begin{bmatrix} \hat{x}_k \\ 0 \end{bmatrix} + \begin{bmatrix} \Delta e_k \\ \Delta e_k \end{bmatrix} + \begin{bmatrix} e_k - \Delta e_k \\ e_k - \Delta e_k \end{bmatrix}.$$

Therefore, $\|z_k\|$ is unbounded if and only if $\|\Delta e_k\|$ is unbounded. The proof is completed. ■

Lemma 1 follows from the stability of the controller and the filter since the stealthiness constraint $D(\bar{r}_k || r_k) \le \delta$ makes $\|\eta_k\|$ bounded. In the following, using Lemma 1, we can study the boundedness of $Rch(z_N, \psi)$ by analyzing that of $Rch(\Delta e_N, 0)$.

A. Finite Time Case

Theorem 1: Under Assumptions 1 and 2, when N is finite, both $Rch(\Delta e_N, 0)$ and $Rch(z_N, \psi))$ are unbounded if and only if $\tilde{D}$ does not have full-column rank.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

Proof: First, we prove the sufficiency. Suppose $\tilde{D}$ does not have full-column rank. Then, there exists a nonzero attack $\xi = \begin{bmatrix}\xi_u^T & \xi_y^T\end{bmatrix}^T$ with $\xi_u \in \mathbf{R}^p$ and $\xi_y \in \mathbf{R}^q$ such that $\tilde{D}\xi = 0$. Since E is injective, we have $\xi_u \neq 0$. From $\mathrm{Null}(\tilde{D}) \subset \mathrm{Null}(\begin{bmatrix}KCD & KE\end{bmatrix})$ and D is injective, it follows that $\bar{B}\xi = D\xi_u \neq 0$. Hence, $Rch(\Delta e_N, 0)$ is unbounded if we choose the attack signals to be $\xi_k = 0$, $k = 0, \dots, N-2$, and $\xi_{N-1} = \lim_{j\to\infty} j\xi$.

Next, we prove the necessity by contradiction. Suppose $\|\Delta e_k\|$ is bounded. Since $\tilde{D}$ has full-column rank, from (11), we know $\|\xi_k\|$ must be bounded to keep $\|\Delta r_{k+1}\| \leq 1$. Then, by (10), we have $\|\Delta e_{k+1}\|$ is bounded. Note that $\|\Delta e_0\| = 0$ is bounded. Hence, $\|\Delta e_N\|$ is bounded by induction. The proof is completed. ■

Intuitively, when $\tilde{D}$ has full-column rank, any nonzero attack ξ_k can change the value of Δr_k. Hence, to keep $\|\Delta r_k\|$ bounded, $\|\xi_k\|$ has to be bounded, which makes $\|\Delta e_N\|$ bounded.

According to Theorem 1, to keep $Rch(z_N, \psi)$ bounded in finite time, we make the following assumption.

Assumption 3: $\tilde{D}$ has full-column rank.

In the sequel, based on Assumption 3, we investigate the boundedness of $Rch(z_N, \psi)$ for the case that $N \to +\infty$.

B. Infinite Time Case

Since $\tilde{D}$ has full-column rank, by (10) and (11), we have

$$\Delta e_{k+1} = \bar{A}\Delta e_k + \bar{K}\Delta r_{k+1} \tag{15}$$

where $\bar{A} = A - \begin{bmatrix}D & 0\end{bmatrix}\tilde{D}^{\dagger}CA$ and $\bar{K} = \begin{bmatrix}D & 0\end{bmatrix}\tilde{D}^{\dagger} - K$.

To derive the necessary and sufficient condition for $Rch(z_N, \psi)$ being unbounded, we need the following lemma.

Lemma 2 [17]: For a matrix $S \in \mathbf{R}^{n\times n}$ and a vector $\beta \in \mathbf{R}^n$, if $\lim_{k\to+\infty} S^k\beta \neq 0$, then there exists an unstable eigenvector γ of S such that $\gamma \in \mathrm{Span}(\begin{bmatrix}\beta & S\beta & \dots & S^{n-1}\beta\end{bmatrix})$.

Then, we obtain the following theorem.

Theorem 2: Under Assumptions 1–3, when $N \to +\infty$, both $Rch(\Delta e_N, 0)$ and $Rch(z_N, \psi)$ are unbounded if and only if $\bar{A}$ has an unstable eigenvalue λ and the corresponding eigenvector ν satisfies the following.

1) $CA\nu \in \mathrm{Span}(\tilde{D})$.

2) ν is a reachable state of (10).

Proof: We first prove the sufficiency.

Since $CA\nu \in \mathrm{Span}(\tilde{D})$, there exists an attack ξ^* such that $\tilde{D}\xi^* = CA\nu$. Moreover, since ν is reachable, there exists an attack sequence $\xi_0^*, \dots, \xi_{n-1}^*$ such that $\Delta e_n = \nu$. Suppose the output of dynamics (10) and (11) with this attack sequence as input are $\Delta r_1^*, \dots, \Delta r_n^*$, and let $\epsilon = \max_{i=1,\dots,n} \|\Delta r_i^*\|$.

If the attack signals are chosen to be

$$\xi_i = \begin{cases} \xi_i^*/\epsilon, & i = 0, \dots, n-1 \\ -\lambda^{i-n}\xi^*/\epsilon, & i \geq n \end{cases}$$

then, by (11) and (15), one has

$$\Delta e_i = \lambda^{i-n}\nu/\epsilon, \quad i \geq n$$
$$\Delta r_i = \begin{cases} \Delta r_i^*/\epsilon, & i = 1, \dots, n \\ 0, & i \geq n+1. \end{cases}$$

Obviously, $\lim_{k\to+\infty} \|\Delta e_k\| = +\infty$ and $\|\Delta r_k\| \leq 1$.

Next, we will prove the necessity.

Since $\lim_{k\to+\infty} \|\Delta e_k\| = +\infty$, there exists a subsequence of $\{\Delta e_i\}$, defined as $\{\Delta e_{i_k}\}$, with $i_0 = 0$ and $i_k = \min\{j \,|\, \|\Delta e_j\| > \|\Delta e_{i_{k-1}}\|\}$. It can be observed that $\|\Delta e_{i_k}\| = +\infty$ as $k \to +\infty$.

Let $g_k = \Delta e_k / \|\Delta e_k\|$. It is obvious that $\|g_k\|$ is bounded. By the Bolzano–Weierstrass theorem and (15), there exists an index set $\{j_k\} \subset \{i_k\}$ such that subsequences $\{g_{j_k}\}$, $\{g_{j_k-1}\}, \dots, \{g_{j_k-n+1}\}$ converge as $k \to +\infty$.

Furthermore, by (15), the following inequality:

$$\|\Delta e_{k+1}\| \leq \|\bar{A}\|\|\Delta e_k\| + \|\bar{K}\|$$

holds. Since $\|\Delta e_{j_k}\| \to +\infty$ as $k \to +\infty$, we have $\lim_{k\to+\infty} \|\Delta e_{j_k-\mu}\| = +\infty$, $\mu = 0, 1, \dots, n-1$.

Define

$$h_\mu = \lim_{k\to+\infty} g_{j_k-\mu}, \quad \mu = 0, 1, \dots, n-1.$$

Then, it follows that:

$$\begin{aligned}\lim_{k\to+\infty} \frac{\Delta e_{j_k-\mu}}{\|\Delta e_{j_k-\mu-1}\|} &= \lim_{k\to+\infty} \frac{\bar{A}\Delta e_{j_k-\mu-1} + \bar{K}\Delta r_{j_k-\mu}}{\|\Delta e_{j_k-\mu-1}\|} \\ &= \bar{A}h_{\mu+1}, \quad \mu = 0, \dots, n-2.\end{aligned} \tag{16}$$

Hence, one has

$$\begin{aligned}h_\mu &= \lim_{k\to+\infty} \frac{\|\Delta e_{j_k-\mu-1}\|}{\|\Delta e_{j_k-\mu}\|} \lim_{k\to+\infty} \frac{\Delta e_{j_k-\mu}}{\|\Delta e_{j_k-\mu-1}\|} \\ &= \frac{\bar{A}h_{\mu+1}}{\|\bar{A}h_{\mu+1}\|}, \quad \mu = 0, \dots, n-2.\end{aligned}$$

Therefore, we obtain

$$\mathrm{Span}(\begin{bmatrix}h_0 & \dots & h_{n-1}\end{bmatrix}) = \mathrm{Span}(\begin{bmatrix}\bar{A}^{n-1}h_{n-1} & \dots & h_{n-1}\end{bmatrix}). \tag{17}$$

From the definition of $\{\Delta e_{i_k}\}$, we have $\|\Delta e_{j_k}\| \geq \|\Delta e_{j_k-1}\|$. Then, from (16), it follows that:

$$\|\bar{A}h_{n-1}\| \geq \lim_{k\to+\infty} \left\| \frac{\Delta e_{j_k-n+1}}{\|\Delta e_{j_k-n+1}\|} \right\| = \|h_{n-1}\|.$$

Hence, $\lim_{k\to+\infty} \bar{A}^k h_{n-1} \neq 0$. By Lemma 2 and (17), there exists an unstable eigenvector ν of $\bar{A}$ such that $\nu \in \mathrm{Span}(\begin{bmatrix}h_0 & \dots & h_{n-1}\end{bmatrix})$.

It follows from (11) and $\|\Delta r_{k+1}\| \leq 1$ that:

$$\left\| \frac{\Delta r_{j_k+1}}{\|\Delta e_{j_k}\|} \right\| = \left\| CAg_{j_k} + \tilde{D}\frac{\xi_k}{\|\Delta e_{j_k}\|} \right\| \leq \frac{1}{\|\Delta e_{j_k}\|}$$

which becomes

$$\left\| CAh_0 + \tilde{D}\frac{\xi_k}{\|\Delta e_{j_k}\|} \right\| \leq 0$$

as $k \to +\infty$. As a result, $CAh_0 \in \mathrm{Span}(\tilde{D})$. Similarly, $CAh_\mu \in \mathrm{Span}(\tilde{D})$ for $\mu = 1, \dots, n-1$. Hence, $CA\nu \in \mathrm{Span}(\begin{bmatrix}CAh_0 & \dots & CAh_{n-1}\end{bmatrix}) \subset \mathrm{Span}(\tilde{D})$.

Moreover, g_k is reachable for the reason that Δe_k is reachable. Since the reachable subspaces are closed, the limits h_μ, $\mu = 0, \dots, n-1$, are reachable and, thus, ν is reachable, which completes the proof. ■

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

Remark 2: The necessary and sufficient condition proposed in [17], where $u_k^a = 0$ and the χ^2 detector rather than the KLD is used, only guarantees unbounded (or bounded by the opposite condition) $Rch(e_N, 0)$ for infinite time case, while our condition is based on Assumption 3, which also guarantees bounded $Rch(e_N, 0)$ for finite time case. Then, in this article, the vector ν is the unstable eigenvector of $\bar{A}$, not that of A in [17]. Hence, we should redesign the attack to prove the sufficiency and cannot obtain the condition $C\nu \in \text{Span}(E)$ in [17] when studying the necessity.

IV. Approximation of the Reachable Set

In this section, we provide a method to approximate the bounded reachable set $Rch(z_N, \psi)$ and evaluate the safety of the system (will be defined in Section IV-B) as an application of this approximation.

A. Approximation Approach

We first introduce the following lemma.

Lemma 3: For vectors $\rho_k \in \mathbf{R}^n$ and $w_k^i \in \mathbf{R}^{n_i}$, with $k \in \mathbf{N}$ and $i = 1, \ldots, \theta$, suppose $(w_k^t)^T W^t w_k^t \leq 1$ and $(w_k^j)^T W_k^j w_k^j \leq \sigma_j$, where $W^t \succ 0$, $W_k^j \succ 0$, and $\sigma_j \geq 0$, for $k \in \mathbf{N}$, $t = 1, \ldots, \theta_0$, and $j = \theta_0 + 1, \ldots, \theta$ with $1 \leq \theta_0 < \theta$. Given a constant $\alpha \in (0, 1)$, if there exist constants $\alpha_k^t \in (0, 1)$, $t = 1, \ldots, \theta_0$, satisfying $\sum_{t=1}^{\theta_0} \alpha_k^t \leq \alpha$ and the function $V : \mathbf{R}^n \mapsto \mathbf{R}_{\geq 0}$ such that the inequality

$$V(\rho_{k+1}) - \alpha V(\rho_k) - \sum_{j=\theta_0+1}^{\theta} \left(w_k^j\right)^T W_k^j w_k^j - \sum_{t=1}^{\theta_0} \alpha_k^t \left(w_k^t\right)^T W^t w_k^t \leq 0, \quad k \in \mathbf{N} \tag{18}$$

holds, then we have

$$V(\rho_k) \leq \alpha^k V(\rho_0) + \left(\alpha + \sum_{j=\theta_0+1}^{\theta} \sigma_j\right) \frac{1 - \alpha^k}{1 - \alpha}. \tag{19}$$

Proof: From (18), it follows that:

$$V(\rho_{k+1}) \leq \alpha V(\rho_k) + \sum_{j=\theta_0+1}^{\theta} \sigma_j + \alpha. \tag{20}$$

Hence, (19) holds by the iteration of (20). ■

From (8) and (9), one has

$$z_{k+1} = \mathcal{A} z_k + B_w w_k + B_v v_{k+1} + B_r r_{k+1} \tag{21}$$

where

$$\mathcal{A} = \begin{bmatrix} A + BL & -BL - \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger CA \\ 0 & A - \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger CA \end{bmatrix}$$

$$B_w = \begin{bmatrix} I_n - \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger C \\ I_n - \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger C \end{bmatrix}, \quad B_v = -\begin{bmatrix} \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger \\ \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger \end{bmatrix}$$

and $B_r = \begin{bmatrix} \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger \\ \begin{bmatrix} D & 0 \end{bmatrix} \tilde{D}^\dagger - K \end{bmatrix}$.

Then, we can derive the following lemma by using Lemma 3.

Lemma 4: Consider (21) with initial state $z_0 = 0$. For a fixed-time horizon N, given a constant $\alpha \in (0, 1)$, if there exist constants α_i, $i = 1, 2, 3$, and matrices $\mathcal{P}$ and Σ that satisfy constraints (23)–(26) in the following convex optimization problem:

Problem 1:

$$\min_{\mathcal{P}, \Sigma, \alpha_1, \alpha_2, \alpha_3} \quad \log\left(\alpha + a + \frac{1}{\alpha_3}\right) - \log|\mathcal{P}| \tag{22}$$

$$\text{s.t.} \quad \alpha_1, \alpha_2 \in (0, 1), \alpha_1 + \alpha_2 \leq \alpha \tag{23}$$

$$\mathcal{P} \succ 0, \quad \Sigma \succ 0, \quad \alpha_3 \geq 1 \tag{24}$$

$$\begin{bmatrix} \alpha \mathcal{P} & 0 & \mathcal{A}^T \mathcal{P} \\ 0 & \mathcal{W} & \mathcal{B}^T \mathcal{P} \\ \mathcal{P} \mathcal{A} & \mathcal{P} \mathcal{B} & \mathcal{P} \end{bmatrix} \succeq 0 \tag{25}$$

$$\text{Tr}(\Sigma Q) - \log(|Q||\Sigma|) - m + \frac{2\delta}{\alpha_3} \leq 2\delta \tag{26}$$

where

$$\mathcal{B} = \begin{bmatrix} B_w & B_v & B_r & B_r \end{bmatrix}$$
$$\mathcal{W} = \text{blkdiag}\left(\alpha_1 \Sigma_w^{-1}/b, \alpha_2 \Sigma_v^{-1}/c, \Sigma, \Sigma/(2\delta)\right)$$

with a, b, and c given in Definition 2, then we have

$$Rch(z_N, 0) \subset \mathcal{E}(z_N, 0) = \left\{ z_N \middle| z_N^T \mathcal{P} z_N \leq \left(\alpha + a + \frac{1}{\alpha_3}\right) \frac{1 - \alpha^N}{1 - \alpha} \right\}.$$

Moreover, let $\alpha_i = \alpha_i^*$, $i = 1, 2, 3$, $\mathcal{P} = \mathcal{P}^*$, and $\Sigma = \Sigma^*$ be the solution of Problem 1, then

$$\mathcal{E}^*(z_N, 0) = \left\{ z_N \middle| z_N^T \mathcal{P}^* z_N \leq \left(\alpha + a + \frac{1}{\alpha_3^*}\right) \frac{1 - \alpha^N}{1 - \alpha} \right\}$$

has the minimum volume among all the ellipsoids $\mathcal{E}(z_N, 0)$ with $\mathcal{P}$ and α_3 satisfying (23)–(26).

Proof: Dynamics (21) can be regarded as a system perturbed by w_k, v_{k+1}, $r_{k+1} - \eta_{k+1}$, and η_{k+1}, which satisfy $w_k^T \Sigma_w^{-1} w_k / b \leq 1$, $v_{k+1}^T \Sigma_v^{-1} v_{k+1} / c \leq 1$, $(r_{k+1} - \eta_{k+1})^T \Sigma_{k+1} (r_{k+1} - \eta_{k+1}) \leq a$, and (14). Furthermore, (14) can be transformed into constraints $\eta_{k+1}^T \Sigma_{k+1} \eta_{k+1} / (2\delta) \leq 1/\alpha_3$, $\alpha_3 \geq 1$, and (26).

In Lemma 3, let $\theta = 4$ and choose the vectors ω_k^i, $i = 1, 2, 3, 4$, and the function $V(\rho_k)$ to be w_k, v_{k+1}, $r_{k+1} - \eta_{k+1}$, η_{k+1}, and $z_k^T \mathcal{P} z_k$, respectively, where $\mathcal{P} \succ 0$. By (18) and (21), one has

$$\bar{z}_k^T \mathcal{Q} \bar{z}_k \geq 0 \tag{27}$$

where $\bar{z}_k^T = \begin{bmatrix} z_k^T & w_k^T & v_{k+1}^T & (r_{k+1} - \eta_{k+1})^T & \eta_{k+1}^T \end{bmatrix}$ and

$$\mathcal{Q} = \begin{bmatrix} \alpha \mathcal{P} - \mathcal{A}^T \mathcal{P} \mathcal{A} & -\mathcal{A}^T \mathcal{P} \mathcal{B} \\ -\mathcal{B}^T \mathcal{P} \mathcal{A} & W - \mathcal{B}^T \mathcal{P} \mathcal{B} \end{bmatrix}.$$

The inequality (27) holds if and only if $\mathcal{Q} \succeq 0$, which is equivalent to the LMI (25) by the Schur complement. Hence, according to Lemma 3, if the constraints in Problem 1 hold, then $Rch(z_N, 0) \subset \mathcal{E}(z_N, 0)$.

Next, we look to minimize the volume of the ellipsoid $\mathcal{E}(z_N, 0)$. For the given constant α and the time horizon N, $(1-\alpha^N)/(1-\alpha)$ is a constant. Then, according to [32], $|\mathcal{P}/(\alpha + a + 1/\alpha_3)|^{-1/2}$ is proportional to the volume of $\mathcal{E}(z_N, 0)$. Since

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

$\log|\mathcal{P}/(\alpha + a + 1/\alpha_3)|^{-1}$ and $|\mathcal{P}/(\alpha + a + 1/\alpha_3)|^{-1/2}$ have the same minimizer, we use $\log|\mathcal{P}/(\alpha + a + 1/\alpha_3)|^{-1}$ as the objective function (22) of Problem 1.

Finally, we prove Problem 1 is convex. The convexity of constraints (23)–(25) is obvious. By [32], the log-concave functions $-\log(|Q||\Sigma|)$ and $-\log|\mathcal{P}|$ are convex. Since $\mathrm{Tr}(\Sigma Q)$ is affine in Σ and the power function $1/\alpha_3$ is convex in $\alpha_3 \geq 1$, constraint (26) is convex. Moreover, the objective function is also convex due to the convexity of $\log(\alpha + a + 1/\alpha_3)$, which completes the proof. ■

Remark 3: Although Problem 1 is convex, the CVX toolbox [33], which is widely used in solving convex optimization problems, is not able to deal with the function $\log(\alpha+a+1/\alpha_3)$ in (22). To make Problem 1 solvable, instead of α_3, we use α_4 such that $1/\alpha_4 = \alpha + a + 1/\alpha_3$ as a variable. Then, the objective function (22) becomes $-\log\alpha_4 - \log|\mathcal{P}|$. Moreover, the α_3 in constraints (24) and (26) should be replaced with $1/(1/\alpha_4 - a - \alpha)$.

Now, we are ready to approximate $Rch(z_N, \psi)$.

Theorem 3: Consider (21) with initial state $z_0 = \psi$. For a fixed-time horizon N, given a constant $\alpha \in (0, 1)$, if there exist constants α_i, $i = 1, 2, 3$, and matrices $\mathcal{P}$ and Σ that satisfy the inequalities (23)–(26), then we obtain

$$\begin{aligned} Rch(z_N, \psi) &\subset \mathcal{E}(z_N, \psi) \\ &= \left\{ z_N \middle| \left(z_N - \mathcal{A}^N\psi\right)^T \mathcal{P}\left(z_N - \mathcal{A}^N\psi\right) \right. \\ &\qquad \left. \leq \left(\alpha + a + \frac{1}{\alpha_3}\right)\frac{1-\alpha^N}{1-\alpha} \right\}. \end{aligned}$$

Moreover, let $\alpha_i = \alpha_i^*$, $i = 1, 2, 3$, $\mathcal{P} = \mathcal{P}^*$, and $\Sigma = \Sigma^*$ be the solution of Problem 1, then

$$\begin{aligned} \mathcal{E}^*(z_N, \psi) = &\left\{ z_N \middle| \left(z_N - \mathcal{A}^N\psi\right)^T \mathcal{P}^*\left(z_N - \mathcal{A}^N\psi\right) \right. \\ &\qquad \left. \leq \left(\alpha + a + \frac{1}{\alpha_3^*}\right)\frac{1-\alpha^N}{1-\alpha} \right\} \end{aligned}$$

has the minimum volume among all the ellipsoids $\mathcal{E}(z_N, \psi)$ with $\mathcal{P}$ and α_3 satisfying (23)–(26).

Proof: It is a direct result from Lemma 4 and the linearity of dynamics (21). ■

From Lemma 4 and Theorem 3, it can be observed that the solution of Problem 1 relies on the pregiven parameter $\alpha \in (0, 1)$. Note that if we also treat α as a variable, Problem 1 is no longer convex. Hence, we use a grid search over (0,1) to find the α that leads to $\mathcal{E}^*(z_N, \psi)$ with minimum volume. Moreover, suppose numbers α_{small} and α_{large} are two choices of α such that $0 < \alpha_{\text{small}} < \alpha_{\text{large}} < 1$. From (23) and (25), we know that if Problem 1 has no solution with $\alpha = \alpha_{\text{large}}$, then it also has no solution with $\alpha = \alpha_{\text{small}}$. Therefore, if the accuracy requirement is not much too high, searching optimal α over a bounded horizon (0, 1) is acceptable.

Remark 4: In [22], another convex optimization problem is constructed to approximate the reachable set for the system suffering stealthy attacks and bounded noises. In this article, the noises are also bounded since in (7), the Gaussian distributions are truncated according to desired probabilities. However, the method proposed in [22] is not applicable anymore. The reason is that the covariance matrix Σ_k and mean value η_k of the innovation r_k are restricted by the KLD. Therefore, to construct Problem 1, we introduce Lemma 3, which is reduced to [22, Lemma 1] when $\sigma_j = 0$, $j = \theta_0 + 1, \ldots, \theta$. Moreover, the additional variable α_3 should be introduced so that Problem 1 is convex.

B. Safety Evaluation

With the approximation $\mathcal{E}^*(z_N, \psi)$ obtained from Theorem 3, we are able to evaluate the safety of the system. The system is safe when $Rch(z_N, \psi)$ and the set of critical states (i.e., z_k that the system is not supposed to reach) have no intersection and vice versa.

We consider two kinds of sets containing critical states which are usually used in many practical applications: the ellipsoid $\mathcal{E} = \{z \in \mathbf{R}^{2n} | (z - \tilde{z})^T \mathcal{S}(z - \tilde{z}) \leq 1\}$ and the union of half-spaces $\mathcal{H} = \{z \in \mathbf{R}^{2n} | \bigcup_{i=1}^{\tau} a_i^T z \geq b_i\}$, where $\tilde{z}$, $a_i \in \mathbf{R}^{2n}$, $b_i \in \mathbf{R}$, $\mathcal{S} \succeq 0$, and τ is a positive integer. For example, to avoid a car running into a stone on the road, we can use $\mathcal{E}$ that describes the location of the stone as the set of critical states. Also, the speed of the car should not be too high, then $\mathcal{H}$ that restricts the speed can be treated as the set of critical states.

Corollary 1: When the set of critical states is $\mathcal{E}$, if inequality $(z_N^* - \tilde{z})^T \mathcal{S}(z_N^* - \tilde{z}) > 1$ holds, then the system is safe at time instant N, where z_N^* is the solution of the following convex optimization problem.

Problem 2:

$$\begin{aligned} \min_{z_N} \quad & (z_N - \tilde{z})^T \mathcal{S}(z_N - \tilde{z}) \\ \text{s.t.} \quad & z_N \in \mathcal{E}^*(z_N, \psi). \end{aligned}$$

When the set of critical states is $\mathcal{H}$, if inequalities

$$\begin{aligned} d_i = &\frac{1}{\sqrt{a_i^T a_i}} \left(|b_i - a_i^T \mathcal{A}^N \psi| - \sqrt{\frac{a_i^T \mathcal{P}^{*-1} a_i (1 - \alpha^N)}{\alpha_4^* - \alpha_4^* \alpha}} \right) \\ &> 0, \quad i = 1, \ldots, \tau \end{aligned}$$

hold, where $1/\alpha_4^* = \alpha + a + 1/\alpha_3^*$, then the system is safe at time instant N.

Proof: From Theorem 3, we know $Rch(z_N, \psi) \subset \mathcal{E}^*(z_N, \psi)$.

For the set $\mathcal{E}$, if the inequality $(z_N - \tilde{z})^T \mathcal{S}(z_N - \tilde{z}) > 1$ holds for all $z_N \in \mathcal{E}^*(z_N, \psi)$, then the system is safe. Hence, the safety of the system is guaranteed when $(z_N^* - \tilde{z})^T \mathcal{S}(z_N^* - \tilde{z}) > 1$. Moreover, the convexity of Problem 2 follows from $\mathcal{P} \succ 0$ and $\mathcal{S} \succeq 0$.

For the set $\mathcal{H}$, from [34], it follows that the minimum distance between the ellipsoid $(z_N - \mathcal{A}^N\psi)^T \mathcal{P}^*(z_N - \mathcal{A}^N\psi) = (1 - \alpha^N)/(\alpha_4^* - \alpha_4^*\alpha)$ and the hyperplane $a_i^T z_N = b_i$ is given by d_i. When $d_i \leq 0$, the hyperplane $a_i^T z_N = b_i$ and the ellipsoid $\mathcal{E}^*(z_N, \psi)$ intersect, which completes the proof. ■

When $\mathcal{E}^*(z_N, \psi)$ and the set of critical states intersect, we may redesign the controller and filter to make the system away from the critical states. However, this may destroy control and filtering performances in practice. Another way is to use a smaller detection threshold δ so that the reachable set has a smaller volume.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

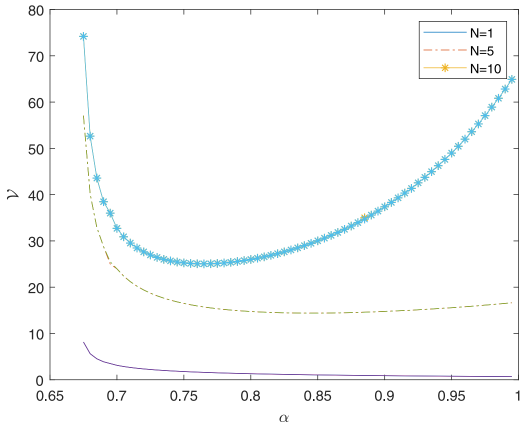


Fig. 1. Relationship between α and the volume of $\mathcal{E}^*(z_N, \psi)$.

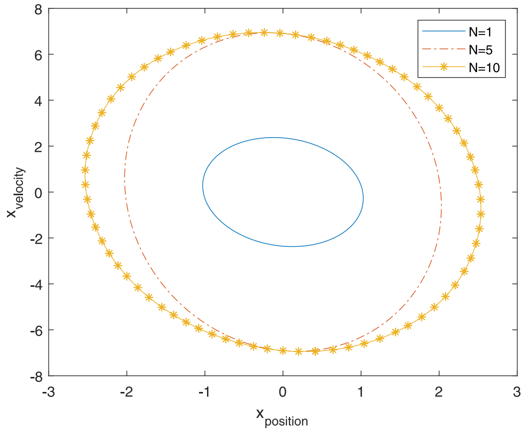


Fig. 2. Projection of $\mathcal{E}^*(z_N, \psi)$ onto the $[x_{\text{position}}, x_{\text{velocity}}]$-hyperplane for different time horizon N.

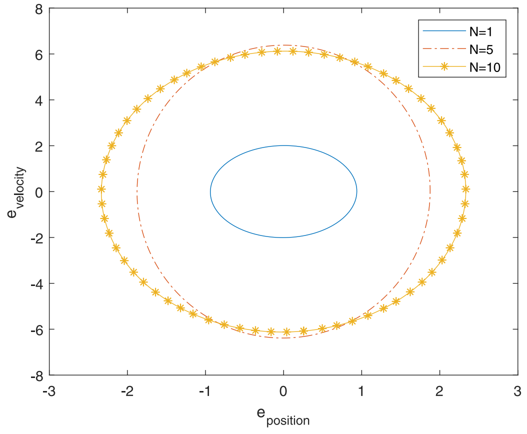


Fig. 3. Projection of $\mathcal{E}^*(z_N, \psi)$ onto the $[e_{\text{position}}, e_{\text{velocity}}]$-hyperplane for different time horizon N.

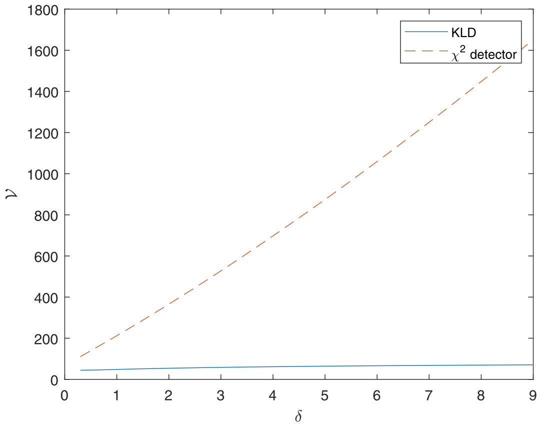


Fig. 4. Volume of the reachable sets with the KLD and the χ^2 detector as the stealthiness constraint, respectively.

It should be pointed out that the word "safe" in Corollary 1 does not mean the system cannot reach the critical states, since the system we consider is disturbed by Gaussian noises. According to $Rch(z_N, \psi) \subset \mathcal{E}^*(z_N, \psi)$ and Definition 2, if the parameters a, b, and c are sufficiently large, and the conditions in Corollary 1 are satisfied, the probability that the system reaches critical states is small.

V. Numerical Example

In this section, the example of an unmanned ground vehicle (UGV) [35] is given to illustrate the effectiveness of the main results. The UGV is assumed to move along a straight line and completely stop in the initial state. Under these assumptions, the system dynamics of the UGV is given by

$$\begin{bmatrix}\dot{x}_{\text{position}}\\ \dot{x}_{\text{velocity}}\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & -\frac{W}{M}\end{bmatrix}\begin{bmatrix}x_{\text{position}}\\ x_{\text{velocity}}\end{bmatrix} + \begin{bmatrix}0\\ \frac{1}{M}\end{bmatrix}u$$

where x_{position}, x_{velocity}, u, $M = 0.5$ kg, and $W = 1$ are the UGV position, its velocity, the force input to the UGV, the mechanical mass, and the translational friction coefficient, respectively.

The model is discretized with 0.1 s, that is, the matrices A and B in (1) are $\begin{bmatrix}1 & 0.0906\\ 0 & 0.8187\end{bmatrix}$ and $\begin{bmatrix}0.0094\\ 0.1813\end{bmatrix}$, respectively. Other matrices in the dynamics (1) and (2) are chosen to be $C = I_2$, $D = \begin{bmatrix}1 & 0\end{bmatrix}^T$, $E = \begin{bmatrix}0 & 1\end{bmatrix}^T$, and $\Sigma_w = \Sigma_v = 0.005I_2$. The controller gain is $L = \begin{bmatrix}-26.4799 & -5.3552\end{bmatrix}$. Both the initial state x_0 and the estimation error e_0 are $\begin{bmatrix}0 & 0\end{bmatrix}^T$. The threshold δ of the KLD equals to 100. The probabilities that the noises and the innovation are within the ellipsoids in Definition 2 are chosen to be 99%, that is, the parameters a, b, and c in (7) equal to 9.21.

For different $\alpha \in (0, 1)$, Problem 1 is solved with the variable α_3 changed by α_4 according to Remark 3. As mentioned in the proof of Lemma 4, $\mathcal{V} = |\mathcal{P}(1-\alpha)\alpha_4/(1-\alpha^N)|^{-(1/2)}$ is proportional to the volume of $\mathcal{E}^*(z_N, \psi)$ [the approximation of $Rch(z_N, \psi)$]. Fig. 1 shows the impact of the parameter $\alpha \in (0.675, 1)$ on the volume of $\mathcal{E}^*(z_N, \psi)$ for $N = 1$, 5, and 10. It should be pointed out that when $\alpha \leq 0.67$, Problem 1 has no solution. We can observe that $\alpha = 0.995$, 0.85, and 0.765 lead to the tightest approximation of $Rch(z_1, \psi)$, $Rch(z_5, \psi)$, and $Rch(z_{10}, \psi)$, respectively.

When α is chosen to be 0.995, 0.85, and 0.765 for $N = 1$, 5, and 10, the projection of $\mathcal{E}^*(z_N, \psi)$ onto the $[x_{\text{position}}, x_{\text{velocity}}]$-hyperplane and $[e_{\text{position}}, e_{\text{velocity}}]$-hyperplane is shown in Figs. 2 and 3, respectively, where $[e_{\text{position}}, e_{\text{velocity}}]$ is the estimate error of the Kalman filter for $[x_{\text{position}}, x_{\text{velocity}}]$. It can be observed that the volume of $\mathcal{E}^*(z_N, \psi)$ grows with the increase of the time horizon N.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

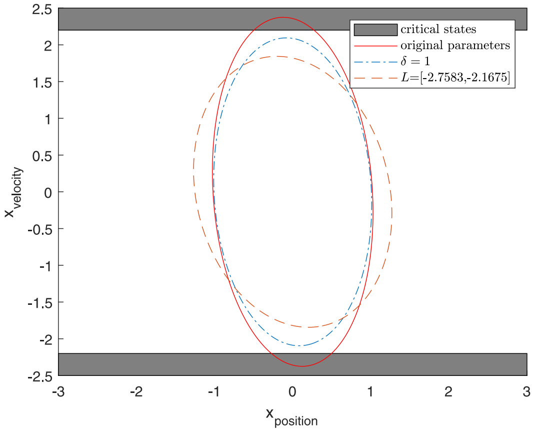


Fig. 5. Projection of $\mathcal{E}^*(z_1, \psi)$ onto the $[x_{\text{position}}, x_{\text{velocity}}]$-hyperplane for different systems parameters.

Next, we compare the reachable sets approximated with the method of [22] and Theorem 3. Note that the energy-based detector considered in [22] becomes the χ^2 detector when the noises are Gaussian. The pregiven parameter α is set to be 0.9 and the time horizon N is chosen to be $+\infty$. The volume of the two sets for different δ is shown in Fig. 4. We can observe that the reachable set approximated with Theorem 3 always has a smaller volume. The reason is that with the KLD adopted, the attacks should be stealthy against not only the χ^2 detector but also other detectors.

Finally, we show how to adjust parameters to make the system safe. Suppose when $N = 1$, the velocity of the UGV is not supposed to be larger than 2.2, that is, the set of critical states is $\mathcal{H} = \{x_{\text{velocity}} > 2.2 \text{ or } x_{\text{velocity}} < -2.2\}$. From Fig. 5, we can observe that the original parameters cannot make the UGV safe. If we change the controller gain L to $\begin{bmatrix}-2.7583 & -2.1675\end{bmatrix}$ or change the threshold δ to 1, the velocity will always be smaller than 2.2. However, since changing L can vary the poles of the closed-loop system, adjusting δ might be a better choice.

VI. Conclusion

In this work, we studied the reachable set of the CPS with the KLD as a detection constraint. We defined the reachable set according to the statistical characteristics of the stochastic system. The necessary and sufficient conditions of the reachable set being unbounded were given for the finite and infinite time cases, respectively. The outer ellipsoid that approximates the bounded reachable set with minimum volume was obtained by solving a convex optimization problem. An application of this approximation to the safety evaluation was presented. Future work involves expanding the results to CPSs with the participant of other kinds of filters or controllers and to large-scale systems.

References

[1] P. Derler, E. A. Lee, and A. S. Vincentelli, "Modeling cyber–physical systems," *Proc. IEEE*, vol. 100, no. 1, pp. 13–28, Jan. 2012.

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

[3] M. Wolf and D. Serpanos, "Safety and security in cyber–physical systems and Internet-of-Things systems," *Proc. IEEE*, vol. 106, no. 1, pp. 9–20, Jan. 2018.

[4] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber–physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, Jan. 2018.

[5] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, "Secure control systems: A quantitative risk management approach," *IEEE Control Syst. Mag.*, vol. 35, no. 1, pp. 24–45, Feb. 2015.

[6] J. Milošević, A. Teixeira, T. Tanaka, K. H. Johansson, and H. Sandberg, "Security measure allocation for industrial control systems: Exploiting systematic search techniques and submodularity," *Int. J. Robust Nonlinear Control*, vol. 30, no. 11, pp. 4278–4302, 2020.

[7] J. Giraldo *et al.*, "A survey of physics-based attack detection in cyber–physical systems," *ACM Comput. Surveys*, vol. 51, no. 4, p. 76, 2018.

[8] E. Mousavinejad, F. Yang, Q.-L. Han, and L. Vlacic, "A novel cyber attack detection method in networked control systems," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 3254–3264, Nov. 2018.

[9] Z.-H. Pang, L.-Z. Fan, J. Sun, K. Liu, and G.-P. Liu, "Detection of stealthy false data injection attacks against networked control systems via active data modification," *Inf. Sci.*, vol. 546, pp. 192–205, Feb. 2021.

[10] H. Yuan, Y. Xia, H. Yang, and Y. Yuan, "Resilient control for wireless networked control systems under DoS attack via a hierarchical game," *Int. J. Robust Nonlinear Control*, vol. 28, no. 15, pp. 4604–4623, 2018.

[11] S. R. Etesami and T. Başar, "Dynamic games in cyber–physical security: An overview," *Dyn. Games Appl.*, vol. 9, pp. 884–913, Jan. 2019.

[12] D. Ding, Z. Wang, D. W. Ho, and G. Wei, "Observer-based event-triggering consensus control for multiagent systems with lossy sensors and cyber-attacks," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 1936–1947, Aug. 2017.

[13] K. Liu, H. Guo, Q. Zhang, and Y. Xia, "Distributed secure filtering for discrete-time systems under Round-Robin protocol and deception attacks," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3571–3580, Aug. 2020.

[14] Y. Chen, S. Kar, and J. M. Moura, "Cyber–physical attacks with control objectives," *IEEE Trans. Autom. Control*, vol. 63, no. 5, pp. 1418–1425, May 2018.

[15] Z.-H. Pang, G.-P. Liu, D. Zhou, F. Hou, and D. Sun, "Two-channel false data injection attacks against output tracking control of networked systems," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3242–3251, May 2016.

[16] 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, Mar. 2018.

[17] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Proc. 1st Workshop Secure Control Syst.*, 2010, pp. 1–6.

[18] C. Kwon, W. Liu, and I. Hwang, "Analysis and design of stealthy cyber attacks on unmanned aerial systems," *J. Aerosp. Inf. Syst.*, vol. 11, no. 8, pp. 525–539, 2014.

[19] I. Jovanov and M. Pajic, "Relaxing integrity requirements for attack-resilient cyber–physical systems," *IEEE Trans. Autom. Control*, vol. 64, no. 12, pp. 4843–4858, Dec. 2019.

[20] Y. Mo and B. Sinopoli, "On the performance degradation of cyber–physical systems under stealthy integrity attacks," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2618–2624, Sep. 2015.

[21] C. Kwon and I. Hwang, "Reachability analysis for safety assurance of cyber–physical systems against cyber attacks," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2272–2279, Jul. 2018.

[22] C. Murguia, I. Shames, J. Ruths, and D. Nešić, "Security metrics and synthesis of secure control systems," *Automatica*, vol. 115, May 2020, Art. no. 108757.

[23] R. Langners. (Nov. 2013). *To Kill a Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve*. [Online]. Available: https://www.langner.com/wp-content/uploads/2017/04/To-kill-a-centrifuge.pdf

[24] 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, Aug. 2017.

[25] C. Fang, J. Chen, Y. Qi, R. Tan, and W. X. Zheng, "Stealthy actuator signal attacks in stochastic control systems: Performance and limitations," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3927–3934, Sep. 2020.

[26] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic Gaussian control systems: Detectability reward tradeoff," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1555–1570, Jul. 2017.

[27] Q. Zhang, K. Liu, Y. Xia, and A. Ma, "Optimal stealthy deception attack against cyber–physical systems," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3963–3972, Sep. 2020.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.

[28] Y.-G. Li and G.-H. Yang, "Optimal stealthy false data injection attacks in cyber–physical systems," *Inf. Sci.*, vol. 481, pp. 474–490, May 2019.

[29] R. Zhang and P. Venkitasubramaniam, "False data injection and detection in LQG systems: A game theoretic approach," *IEEE Trans. Control Netw. Syst.*, vol. 7, no. 1, pp. 338–348, Mar. 2020.

[30] S. Kullback, *Information Theory and Statistics*. North Chelmsford, MA, USA: Courier Corp., 1997.

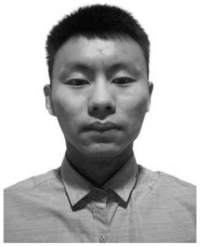
[31] Z. Guo, D. Shi, D. E. Quevedo, and L. Shi, "Secure state estimation against integrity attacks: A Gaussian mixture model approach," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 194–207, Jan. 2019.

[32] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA, USA: SIAM, 1994.

[33] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1.* [Online]. Available: http://cvxr.com/cvx

[34] A. A. Kurzhanskiy and P. Varaiya, "Ellipsoidal toolbox (ET)," in *Proc. 45th Conf. Decis. Control*, San Diego, CA, USA, 2006, pp. 1498–1503.

[35] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber–physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 4917–4932, Oct. 2017.



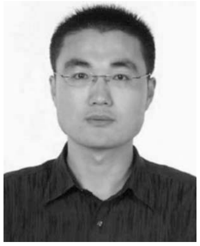
Qirui Zhang received the B.Eng. degree in automation from the Beijing Institute of Technology, Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Automation.

His current research interests include optimal control and the security of cyber-physical systems.

Kun Liu (Member, IEEE) received the Ph.D. degree in electrical engineering and systems from Tel Aviv University, Tel Aviv-Yafo, Israel, in 2012.

From 2013 to 2015, he was a Postdoctoral Researcher with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. In 2015, he held Researcher, Visiting, and Research Associate positions with, respectively, the KTH Royal Institute of Technology; CNRS, Laboratory for Analysis and Architecture of Systems, Toulouse, France; and the University of Hong Kong, Hong Kong. In 2018, he was a Visiting Scholar with INRIA, Lille, France. In 2015, he joined the School of Automation, Beijing Institute of Technology, Beijing, China, where he is currently a tenured Associate Professor. His current research interests include networked control, game-theoretic control, and security and privacy of cyber-physical systems, with applications in autonomous systems.

Dr. Liu currently serves as an Associate Editor for the *IMA Journal of Mathematical Control and Information* and the *Journal of Beijing Institute of Technology*. He is a Conference Editorial Board Member of the IEEE Control Systems Society.



Zhonghua Pang (Senior Member, IEEE) received the B.Eng. degree in automation and the M.Eng. degree in control theory and control engineering from the Qingdao University of Science and Technology, Qingdao, China, in 2002 and 2005, respectively, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011.

He was a Postdoctoral Fellow with the Department of Automation, Tsinghua University, Beijing, from 2011 to 2014. He is currently a Professor with the School of Electrical and Control Engineering, North China University of Technology, Beijing. His research interests include networked control systems, security of cyber-physical systems, and advanced control in the environmental protection industry.

Yuanqing Xia (Senior Member, IEEE) received the M.Sc. degree in fundamental mathematics from Anhui University, Hefei, China, in 1998, and the Ph.D. degree in control theory and control engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2001.

From 2002 to 2003, he was a Postdoctoral Research Associate with the Institute of Systems Science, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, Beijing, where he worked on navigation, guidance, and control. From 2003 to 2004, he was with the National University of Singapore, Singapore, as a Research Fellow, where he worked on variable structure control. From 2004 to 2006, he was with the University of Glamorgan, Pontypridd, U.K., as a Research Fellow, where he studied networked control systems. From 2007 to 2008, he was a Guest Professor with Innsbruck Medical University, Innsbruck, Austria, where he worked on biomedical signal processing. Since 2004, he has been with the School of Automation, Beijing Institute of Technology, Beijing, first as an Associate Professor and then since 2008 as a Professor. His research interests include cloud control systems, networked control systems, robust control and signal processing, active disturbance rejection control, unmanned system control, and flight control.

Prof. Xia is a Deputy Editor of the *Journal of Beijing Institute of Technology* and an Associate Editor of *Acta Automatica Sinica*, *Control Theory and Applications*, the *International Journal of Innovative Computing, Information, and Control*, and the *International Journal of Automation and Computing*.

Tao Liu received the Ph.D. degree in control science and engineering from the University of Science and Technology Beijing, Beijing, China, in 2011.

He is currently a Lecturer with the School of Information, Beijing Wuzi University, Beijing. His research interests include networked control systems, discrete-time system, and sliding-mode control.

Authorized licensed use limited to: ZheJiang Sci-Tech University. Downloaded on October 23,2022 at 11:26:22 UTC from IEEE Xplore. Restrictions apply.