On the Performance Degradation of Cyber-Physical Systems under Stealthy Integrity Attacks

Yilin Mo*, Bruno Sinopoli†

***Abstract*—This paper analyzes the effect of stealthy integrity attacks on Cyber-Physical Systems, which is modeled as a Stochastic Linear Time-Invariant (LTI) system equipped with a linear filter, a linear feedback controller and a χ^2 failure detector. An attacker wishes to induce perturbation in the control loop by compromising a subset of the sensors and injecting an exogenous control input, without incurring detection from an anomaly detector. We show how the problem can be modeled, from the attacker's standpoint, as a constrained control problem and that the characterization of the maximum perturbation can be posed as reachable set computation, which we solve using ellipsoidal calculus.**

I. Introduction

Cyber-Physical Systems (CPS) refer to the embedding of widespread sensing, networking, computation and control into physical spaces with the goal of making them safer, more efficient and reliable. Driven by the miniaturization and integration of sensing, communication and computation in cost effective devices, CPSs are bound to transform many industries such as aerospace, transportation, built environments, energy, health-care, and manufacturing, to name a few. However, using off-the-shelf networking and computing devices provides several opportunities for malicious entities to inject attacks on CPS. A wide variety of motivations exists for launching an attack on the CPSs, ranging from financial reasons, i.e. draw a financial gain, all the way to terrorism, e.g., threatening the life of possibly an entire population by controlling electricity and other life-critical resources. Any successful attack on safety-critical CPSs may significantly hamper the economy, and even lead to the loss of human lives. While the threat of attacks on CPS tend to be underplayed at times, more recently Stuxnet [1] provided a clear sample of the future to come. The research community has acknowledged the importance of addressing the challenge of designing secure CPS [2].

Classical system theory based approaches, such as robust statistic [3] and robust control [4], seek to design algorithms which can withstand certain types of failures. In addition to robust method, Fault Detection and Isolation (FDI) have been extensively studied over the past decades [5]. The main drawback of such an approach is that the failures are usually assumed to be benign, independent or random, while an attack could be carefully designed to exploit certain vulnerabilities of the system. Therefore, the applicability of robust and FDI techniques needs to be carefully reexamined when dealing with CPS security.

In the context of dynamical systems, Pasqualetti et al. [6], [7], Sundaram et al. [8] and Fawzi et al. [9] show how to detect and identify malicious behaviors in consensus networks, power grids, wireless control networks and control systems. However, in the majority of these contributions, the system model is assumed to be noiseless, which greatly favors the failure detector, since the evolution of the system is deterministic and any deviation from the predetermined trajectory will be detected. As a consequence, in all the above papers, the attacker can either arbitrarily perturb the system along certain directions or cannot induce any perturbation, without incurring detection. We believe that a more realistic scenario needs to account for a noisy environment. In this case it is harder to detect malicious behavior since the adversary may inject an attack which inflicts a large perturbation on the system state, while only causing a slightly increasing in the detection rate.

In this paper we focus on developing tools to quantify the maximum perturbation that an attacker can introduce into a control system via a stealthy integrity attack on a subset of the sensors and through the injection of exogenous control inputs. The system is modeled as a Stochastic Linear Time-Invariant (LTI) system equipped with a linear filter, a linear feedback controller and a χ^2 failure detector. We formulate the attacker's action as a constrained control problem and quantify the resilience of the CPS against such attacks using the concept of invariant and reachable set. We further provide a recursive algorithm to compute the inner and outer approximation of the reachable set of the attacker, thus providing a computational method to quantify the maximum perturbation inflicted by a stealthy attack. This article generalizes the preliminary results in [10], [11], where we consider attacks on the sensors only.

The rest of the paper is organized as follows: In Section II, we describe the physical system model and the cyber system model (the estimator, controller and failure detector) of the CPS. In Section III, we introduce the stealthy integrity attack model and formulate the attacker's strategy as a constrained control problem. In Section IV, we design an ellipsoidal approximation algorithm to compute the reachable region of the system under stealthy integrity attacks. In Section V, we provide a numerical example to illustrate the effect of stealthy integrity attacks on CPS. Finally, Section VI concludes the paper.

Notations: $\mathbb{S}_+^n$ is the set of $n \times n$ positive semidefinite matrices. We write $X \geq Y$ if $X - Y \in \mathbb{S}_+^n$. X^+ is the Moore-Penrose pseudoinverse of the matrix X. $\|\cdot\|$ denotes the 2-norm of a vector.

II. System Description

We model the system as a linear control system, which is equipped with a linear filter, a linear feedback controller and a χ^2 failure detector. We assume that the physical system follows:

$$x(k+1) = Ax(k) + Bu(k) + w(k), \tag{1}$$

where $x(k) \in \mathbb{R}^n$ is the vector of state variables at time k, $u(k) \in \mathbb{R}^p$ is the control input, $w(k) \in \mathbb{R}^n$ is the process noise at time k and $x(0)$ is the initial state. $w(k)$, $x(0)$ are independent Gaussian random variables, and $x(0) \sim \mathcal{N}(0,\ \Sigma)$, $w(k) \sim \mathcal{N}(0,\ Q)$.

A sensor network is deployed to monitor the system described in (1). At each step all the sensor readings are collected and sent to a centralized estimator. The observation equation can be written as

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

where $y(k) = [y_1(k), \ldots, y_m(k)]^T \in \mathbb{R}^m$ is a vector of sensor measurements, and $y_i(k)$ is the measurement made by sensor i at time k. $v(k) \sim \mathcal{N}(0,\ R)$ is i.i.d. measurement noise independent of $x(0)$ and $w(k)$.

A linear filter is used to compute state estimation $\hat{x}(k)$ from observations $y(k)$:

$$\begin{aligned}\hat{x}(k+1) = A\hat{x}(k) + Bu(k) + \\ K\left\{y(k+1) - C\left[A\hat{x}(k) + Bu(k)\right]\right\}.\end{aligned} \tag{3}$$

Define the residue $z(k)$ and the estimation error $e(k)$ at time k as

$$z(k) \triangleq y(k) - C(A\hat{x}(k) + Bu(k)),\ e(k) \triangleq x(k) - \hat{x}(k). \tag{4}$$

This research is supported in part by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office Foundation and grant NGIT2009100109 from Northrop Grumman Information Technology Inc Cybersecurity Consortium. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, or the U.S. Government or any of its agencies.

*: Control and Dynamical Systems, California Institute of Technology, Pasadena, CA. Email: yilinmo@caltech.edu

†: Dept of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA. Email: brunos@ece.cmu.edu

We assume that an LTI feedback controller is used to stabilize the system, which takes the following form:

$$u(k) = L\hat{x}(k) \tag{5}$$

It is well known that the closed-loop system is stable if and only if both $A - KCA$ and $A + BL$ are stable [12]. For the rest of the discussion we only focus on systems that are closed-loop stable and in steady state.

Consider the CPS consisting of the physical system, the linear filter and controller. We can immediately identify $x(k)$ as the "physical" state and $\hat{x}(k)$ as the "cyber" state. Thus, we define the state of the system $\tilde{x}(k)$ as:

$$\tilde{x}(k) \triangleq \begin{bmatrix} x(k) \\ e(k) \end{bmatrix} = \begin{bmatrix} I_n & 0 \\ I_n & -I_n \end{bmatrix} \begin{bmatrix} x(k) \\ \hat{x}(k) \end{bmatrix} \in \mathbb{R}^{2n} \tag{6}$$

A. χ^2 Failure Detector

Failure detectors are often used to detect anomalous operations. We assume that a χ^2 failure detector ([13],[14]) is deployed, which computes the following quantity

$$g(k) = z(k)^T P_z^{-1} z(k), \tag{7}$$

where P_z is the covariance matrix of the residue $z(k)$ is a constant matrix since we assume the system is in steady state. Define $P_z^{-\frac{1}{2}}$ to be a symmetric matrix such that $P_z^{-\frac{1}{2}} \times P_z^{-\frac{1}{2}} = P_z^{-1}$. Thus, (7) can be rewritten as $g(k) = \|P_z^{-\frac{1}{2}} z(k)\|^2$.

Since $z(k)$ is Gaussian distributed [13], $g(k)$ is χ^2 distributed with m degrees of freedom. The χ^2 failure detector compares $g(k)$ with a threshold η and triggers ad alarm if $g(k)$ is greater than η. Let us define the probability of triggering an alarm at time k as

$$\beta(k) \triangleq P(g(k) \geq \eta). \tag{8}$$

When the system is operating normally, $\beta(k)$ is a constant, which is defined as the false alarm rate α. In common practice, α is small since false alarms tend to increase operation cost.

III. Threat Model

In this section we describe the integrity attack model on the CPS. We assume that an adversary has the following capabilities:

1) The adversary knows the static parameters of the system, namely A, B, C, K, L, Q, R matrices.
2) The adversary compromises a subset $\{i_1, \ldots, i_l\} \subseteq \{1, \ldots, m\}$ of sensors. The adversary can add arbitrary bias to the readings of the compromised sensors. Define the sensor selection matrix Γ as

$$\Gamma \triangleq [e_{i_1}, \ldots, e_{i_l}] \in \mathbb{R}^{m \times l}, \tag{9}$$

where e_i is the ith vector of the canonical basis of $\mathbb{R}^m$. Further define the bias injected by the attacker $y^a(k) \in \mathbb{R}^l$ as

$$y^a(k) \triangleq [y^a_{i_1}(k), \ldots, y^a_{i_l}(k)]^T,$$

where $y^a_i(k)$ indicates the injected bias on sensor i at time k. Thus, the modified reading received by the estimator can be written as

$$y(k) = Cx(k) + \Gamma y^a(k) + v(k), \tag{10}$$

3) The adversary can inject external control inputs to the system. As a result, the system equation becomes

$$x(k+1) = Ax(k) + Bu(k) + B^a u^a(k) + w(k), \tag{11}$$

where $B^a \in \mathbb{R}^{n \times q}$ characterizes the direction of control inputs the attacker can inject to the system.
4) Without loss of generality, we assume that the injection of control inputs starts at time 0 and the manipulation of sensor measurements starts at time 1. In other words, $u^a(k) = 0$, for all $k \leq -1$, and $y^a(k) = 0$, for all $k \leq 0$.

To simplify notations, let us define the following matrices:

$$\tilde{A} \triangleq \begin{bmatrix} A + BL & -BL \\ 0 & A - KCA \end{bmatrix} \in \mathbb{R}^{2n \times 2n}, \tag{12}$$

$$\tilde{B} \triangleq \begin{bmatrix} B^a & 0 \\ B^a - KCB^a & -K\Gamma \end{bmatrix} \in \mathbb{R}^{2n \times (q+l)}, \tag{13}$$

$$\tilde{C} \triangleq P_z^{-\frac{1}{2}} \begin{bmatrix} 0 & CA \end{bmatrix} \in \mathbb{R}^{m \times 2n}, \tag{14}$$

$$\tilde{D} \triangleq P_z^{-\frac{1}{2}} \begin{bmatrix} CB^a & \Gamma \end{bmatrix} \in \mathbb{R}^{m \times (q+l)}. \tag{15}$$

and the attacker's input $\zeta(k)$ at time k as $\zeta(k) \triangleq \begin{bmatrix} u^a(k) \\ y^a(k+1) \end{bmatrix}$.

Since the system is linear, the cyber-physical state $\tilde{x}(k)$ can be seen as the sum of two signals: $\tilde{x}^n(k)$, the state generated by noise and $\tilde{x}^c(k)$, the state generated by the attacker's action. Similarly, the residue vector $z(k)$ can be seen as the sum of $z^c(k)$ and $z^n(k)$. One can verify that

$$\tilde{x}^c(k+1) = \tilde{A}\tilde{x}^c(k) + \tilde{B}\zeta(k), \tag{16}$$

$$P_z^{-\frac{1}{2}} z^c(k+1) = \tilde{C}\tilde{x}^c(k) + \tilde{D}\zeta(k), \tag{17}$$

$$\tilde{x}^n(k+1) = \tilde{A}\tilde{x}^n(k) + \begin{bmatrix} I & 0 \\ I - KC & -K \end{bmatrix} \begin{bmatrix} w(k) \\ v(k+1) \end{bmatrix}.$$

$$P_z^{-\frac{1}{2}} z^n(k+1) = \tilde{C}\tilde{x}^n(k) + P_z^{-\frac{1}{2}} \begin{bmatrix} C & I_m \end{bmatrix} \begin{bmatrix} w(k) \\ v(k+1) \end{bmatrix}.$$

We further define the attacker's action $\zeta^\infty \triangleq (\zeta(0), \zeta(1), \ldots)$ as an infinite sequence[1] of $\zeta(k)$s. It is clear that $\tilde{x}^c(k)$ and $z^c(k)$ are functions of ζ^∞. Thus, we can write them as $\tilde{x}^c(k, \zeta^\infty)$ and $z^c(k, \zeta^\infty)$ respectively. For simplicity, we will omit ζ^∞ when there is no confusion.

Our goal is to characterize the evolution of the state $\tilde{x}(k)$ during the integrity attack. It is easy to verify that $\tilde{x}^n(k)$ is a stationary Gaussian process, which has the same statistics as $\tilde{x}(k)$ in the absence of the attacker. Consequently we focus on $\tilde{x}^c(k)$, i.e., the state generated by the attacker's action.

It is clear that without any constraint on the attacker's action, the reachable region of $\tilde{x}^c(k)$ is the reachable subspace of $(\tilde{A}, \tilde{B})$. However, if the adversary does not design its input $\zeta(k)$ cautiously, an alarm may be triggered and the attack may be stopped by the system operator before the attacker achieves its goal. As a result, we restrict our attention to "stealthy" attacks.

In this paper, we assume that attacker constrains its action ζ^∞ to satisfy the following inequality:

$$\|P_z^{-\frac{1}{2}} z^c(k+1)\| = \|\tilde{C}\tilde{x}^c(k) + \tilde{D}\zeta(k)\| \leq \delta. \forall k = 0, 1, \ldots \tag{18}$$

where δ is a design parameter of the attacker. Since $z(k) = z^c(k) + z^n(k)$ and $z^n(k)$ has the same distribution as $z(k)$ in the absence of the attack, the adversary can make $z(k)$ very similar to the "nominal" $z(k)$ by enforcing that $z^c(k)$ is small. In other words, the failure detector can hardly distinguish a system that is under attack from a "healthy" system. Such an observation is formalized by the following theorem:

Theorem 1. *For any $\delta \in (0, \sqrt{\eta})$, if (18) holds for all k, then*

$$\beta(k) \leq (\Gamma(m/2))^{-1} \Gamma(m/2, (\sqrt{\eta} - \delta)^2/2)), \tag{19}$$

[1]If the attack stops at time T, then $\zeta(k) = 0$ for all $k > T$.

where $\Gamma(s,x) \triangleq \int_x^\infty t^{s-1}e^{-t}dt$ *is the upper incomplete gamma function and* $\Gamma(s) \triangleq \Gamma(s,0)$ *is the gamma function. Furthermore,*

$$\lim_{\delta\to 0^+} \left(\Gamma(m/2)\right)^{-1}\Gamma(m/2,(\sqrt{\eta}-\delta)^2/2)) = \alpha. \tag{20}$$

Proof. By triangle inequality, we know that

$$\begin{aligned} g(k) &= \|P_z^{-1/2}z^n(k) + P_z^{-1/2}z^c(k)\|^2 \\ &\le \left(\|P_z^{-1/2}z^n(k)\| + \|P_z^{-1/2}z^c(k)\|\right)^2 \end{aligned}$$

Hence, $g(k) \le \eta$ when $\|P_z^{-1/2}z^n(k)\| \le \sqrt{\eta}-\delta$, which implies that

$$\beta(k) \le P\left(\|P_z^{-1/2}z^n(k)\|^2 \le (\sqrt{\eta}-\delta)^2\right). \tag{21}$$

By the properties of χ^2 distribution [15], the RHS of (21) equals the RHS of (19). Furthermore, $\Gamma(s,x)$ is continuous with respect to x by its definition, hence (20) holds. □

As a consequence of (18), we can model the attacker's strategy as a constrained control problem, where the system equation is given by:

$$\tilde{x}^c(k+1) = \tilde{A}\tilde{x}^c(k) + \tilde{B}\zeta(k), \tag{22}$$

and the constraint is as follows:

$$\|\tilde{C}\tilde{x}^c(k) + \tilde{D}\zeta(k)\| \le \delta,\ \forall k = 0,1,\dots \tag{23}$$

Our goal is to compute the reachable region of the state $\tilde{x}^c(k)$, which indicates the resilience of the system against integrity attacks. Due to the linearity of the system, we assume, without loss of generality, that $\delta = 1$ for the rest of the paper, leading to the following definitions:

Definition 1. *The attacker's action* ζ^∞ *is called feasible if* (18) *holds for all* k *and* $\delta = 1$.

Definition 2. *The reachable region* $\mathcal{R}_k$ *of* $\tilde{x}^c(k)$ *is defined as*[2]

$$\mathcal{R}_k \triangleq \{\tilde{x}^c \in \mathbb{R}^{2n} : \tilde{x}^c = \tilde{x}^c(k,\zeta^\infty),\ \textit{for some feasible } \zeta^\infty\}. \tag{24}$$

The union of all $\mathcal{R}_k$ *is defined as:*

$$\mathcal{R} \triangleq \bigcup_{k=0}^{\infty} \mathcal{R}_k. \tag{25}$$

Thus, $\mathcal{R}$ indicates all possible biases that an attacker can inflict into the system.

Remark 1. *For a noiseless system model considered in [6], [7], [8], [9], the adversary has to enforce that* (18) *holds for* $\delta = 0$ *to avoid being detected, as even a small deviation from the nominal behavior of the system will result in an alarm. However, as is illustrated in Section V, it is entirely possible that the attacker, although cannot inject anything when enforcing* $\delta = 0$*, can inflict a large perturbation into the system with a small* δ*, which is hardly detectable in a noisy system. In [6], [7], [8], [9], such an attack with a non-zero* δ *would be considered as a failed attack for the deterministic settings. In this paper,* $\mathcal{R}$ *is used to quantify the performance degradation of a noisy system under the attack.*

IV. Main results

In this section, we consider the problem of computing the reachable set $\mathcal{R}_k$ and $\mathcal{R}$. In Section IV-A, we provide a recursive definition of $\mathcal{R}_k$ based on the concept of controlled invariant set. We further devote Section IV-B to the numerical approximation of $\mathcal{R}_k$ and $\mathcal{R}$.

[2]Notice that the definition of $\mathcal{R}_k$ is different from the definition in [10].

A. Recursive Definition of R_k

Before continuing on, we need to introduce the concept of reach set and one-step set:

Definition 3. *Define the reach set* $\mathrm{Rch}(\mathcal{S})$ *of set* $\mathcal{S} \subseteq \mathbb{R}^{2n}$ *to be*

$$\begin{aligned} \mathrm{Rch}(\mathcal{S}) \triangleq \{\tilde{x}^+ \in \mathbb{R}^{2n} : &\exists \zeta \in \mathbb{R}^{p+l}, \tilde{x}^c \in \mathcal{S}, \\ &s.t., \tilde{A}\tilde{x}^c + \tilde{B}\zeta = \tilde{x}^+,\ \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \le 1\}. \end{aligned} \tag{26}$$

Definition 4. *Define the one-step set* $\mathrm{Pre}(\mathcal{S})$ *of set* $\mathcal{S} \subseteq \mathbb{R}^{2n}$ *to be*

$$\begin{aligned} \mathrm{Pre}(\mathcal{S}) \triangleq \{\tilde{x}^c \in \mathbb{R}^{2n} : &\exists \zeta \in \mathbb{R}^{p+l}, \\ &s.t., \tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{S},\ \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \le 1\}. \end{aligned} \tag{27}$$

Remark 2. *The reach set of* $\mathcal{S}$ *indicates all the states* $\tilde{x}^c(k+1)$ *that can be reached with a one-step admissible attacker's input* $\zeta(k)$*, when the current state* $\tilde{x}^c(k)$ *is in* $\mathcal{S}$*. On the other hand, the one-step set of* $\mathcal{S}$ *indicates all the previous states* $\tilde{x}^c(k-1)$ *that can be driven into* $\mathcal{S}$ *with a one-step admissible* $\zeta(k-1)$.

At the first glance, it seems that $\mathcal{R}_k$ can be recursively defined as $\mathcal{R}_{k+1} = \mathrm{Rch}(\mathcal{R}_k)$. However, the reach set only guarantees that $\|\tilde{C}\tilde{x}^c(k) + \tilde{D}\tilde{\zeta}(k)\| \le 1$ for the current k, not for the future ks. To define $\mathcal{R}_k$ recursively, we need to introduce the concept of controlled invariant set.

Definition 5. *A set* $\mathcal{C} \subseteq \mathbb{R}^{2n}$ *is called controlled invariant if for all* $\tilde{x}^c \in \mathcal{C}$*, there exists a* ζ*, such that the following inequalities hold:*

$$\tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{C},\ \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \le 1. \tag{28}$$

In other words, if the current state $\tilde{x}^c(k)$ belongs to $\mathcal{C}$, then the attacker can always use a admissible $\zeta(k)$ to enforce that the next state $\tilde{x}^c(k+1)$ and hence all the future states to be in $\mathcal{C}$. The following proposition characterizes several important properties of the reach set, the one-step set and the controlled invariant set:

Proposition 1. *The following statements hold for the operator* Pre, Rch *and the controlled invariant set:*

1) Pre *and* Rch *are monotonically nondecreasing, i.e., if* $\mathcal{S}_1 \subseteq \mathcal{S}_2$*, then*
$$\mathrm{Pre}(\mathcal{S}_1) \subseteq \mathrm{Pre}(\mathcal{S}_2),\ \mathrm{Rch}(\mathcal{S}_1) \subseteq \mathrm{Rch}(\mathcal{S}_2). \tag{29}$$
2) *Let* $\mathcal{C}$ *to be a controlled invariant set, then* $\mathcal{C} \subseteq \mathrm{Pre}(\mathcal{C})$.
3) *There exists the maximum controlled invariant set* $\mathcal{C}_\infty$*, such that* $\mathcal{C} \subseteq \mathcal{C}_\infty$ *for all controlled invariant set* $\mathcal{C}$.
4) *Let* $\mathcal{C}_0 = \mathbb{R}^{2n}$ *and* $\mathcal{C}_{k+1} = \mathrm{Pre}(\mathcal{C}_k)$*. Then the following equality holds:*
$$\mathcal{C}_\infty = \bigcap_{k=0}^{\infty} \mathcal{C}_k. \tag{30}$$

Proof. The proof of the first three properties can be found in [16], while the proof of the last property is quite technical and is reported in the appendix to improve legibility[3]. □

Remark 3. *Notice that Proposition 4 in [17] cannot be used to prove the last statement of Proposition 1, since it requires compactness, which may not hold in our case.*

We are now ready to provide a recursive definition of $\mathcal{R}_k$:

Theorem 2. $\mathcal{R}$ *is controlled invariant, and hence* $\mathcal{R} \subseteq \mathcal{C}_\infty$*. Furthermore* $\mathcal{R}_k$ *satisfies the following recursive equation*

$$\mathcal{R}_{k+1} = \mathrm{Rch}(\mathcal{R}_k) \bigcap \mathcal{C}_\infty,\ \textit{with } \mathcal{R}_0 = \{0\}. \tag{31}$$

[3]It is worth noticing that for general systems and feasibility constraints, (30) is not necessarily true[17].

Proof. First we need to prove that $\mathcal{R}$ is controlled invariant. By definition, for any $\tilde{x}^c \in \mathcal{R}$, there exists k and a feasible ζ^∞, such that $\tilde{x}^c = \tilde{x}^c(k, \zeta^\infty)$. As a result, $\zeta(k)$ is the admissible control input to ensure that (28) holds, which implies that $\mathcal{R}$ is controlled invariant. Thus, $\mathcal{R} \subseteq \mathcal{C}_\infty$ due to the maximality of $\mathcal{C}_\infty$.

We now prove (31) by induction. Since the attack starts at time 0, $\mathcal{R}_0 = \{0\}$. Now assume that (31) holds for k. From the definition of $\mathcal{R}_{k+1}$, and the fact that $\mathcal{R}_{k+1} \subseteq \mathcal{R} \subseteq \mathcal{C}_\infty$, it is trivial to prove that $\mathcal{R}_{k+1} \subseteq \mathrm{Rch}(\mathcal{R}_k) \bigcap \mathcal{C}_\infty$. Therefore, we only need to prove the opposite side of the set inclusion, i.e., for all $\tilde{x}^c \in \mathrm{Rch}(\mathcal{R}_k) \bigcap \mathcal{C}_\infty$, there exists a feasible ζ^∞ that drives state $\tilde{x}^c(k+1, \zeta^\infty)$ at time $k+1$ to $\tilde{x}^c$. We construct such ζ^∞ in three steps:

1) By the fact that $\tilde{x}^c \in \mathrm{Rch}(\mathcal{R}_k)$, we know that there exists an $\tilde{x}^c(k) \in \mathcal{R}_k$ and $\zeta(k)$, such that
$$\tilde{x}^c = \tilde{A}\tilde{x}^c(k) + \tilde{B}\zeta(k), \|\tilde{C}\tilde{x}^c(k) + \tilde{D}\zeta(k)\| \leq 1.$$
2) Since $\tilde{x}^c(k) \in \mathcal{R}_k$, by the induction assumption, we know that there exist $\zeta(0), \ldots, \zeta(k-1)$ and $\tilde{x}^c(0) = 0, \ldots, \tilde{x}^c(k-1)$, such that for all $t = 0, \ldots, k-1$.
$$\tilde{x}^c(t+1) = \tilde{A}\tilde{x}^c(t) + \tilde{B}\zeta(t), \|\tilde{C}\tilde{x}^c(t) + \tilde{D}\zeta(t)\| \leq 1.$$
3) Since $\tilde{x}^c \in \mathcal{C}_\infty$, one can find an admissible control $\zeta(k+1)$, such that (28) holds. Now since $\tilde{x}^c(k+2) = \tilde{A}\tilde{x}^c + \tilde{B}\zeta(k+1)$ also belongs to $\mathcal{C}_\infty$, we can repeat the procedure above to find $\zeta(k+2)$, $\zeta(k+3), \ldots$, to ensure (28) holds for all k.

Therefore, $\zeta^\infty = (\zeta(0), \ldots, \zeta(k), \ldots)$ is the required feasible sequence, which concludes the proof. □

Proposition 1 and Theorem 2 enable the computation of $\mathcal{C}_k$ and $\mathcal{R}_k$ by recursively applying the operator Pre and Rch. However, computing the exact shapes of these sets is numerically intractable as k goes to infinity. One standard technique to attack this problem is to compute the inner and outer approximation of $\mathcal{C}_k$ and $\mathcal{R}_k$, using ellipsoids or polytopes. In this paper, we use an ellipsoidal approximation procedure similar to the one proposed in [18]. The detailed approach is presented in the next subsection.

B. Ellipsoidal Approximation of R_k

This section is devoted to constructing an ellipsoidal inner and outer approximation of $\mathcal{C}_k$ and $\mathcal{R}_k$. To this end, let us assume that $\mathcal{C}_k$ and $\mathcal{R}_k$ are approximated by the following ellipsoids:

$$\begin{aligned} \mathcal{E}_{2n}(C^{in}(k)) \subseteq \mathcal{C}_k \subseteq \mathcal{E}_{2n}(C^{out}(k)), \\ \mathcal{E}_{2n}(R^{in}(k)) \subseteq \mathcal{R}_k \subseteq \mathcal{E}_{2n}(R^{out}(k)), \end{aligned} \tag{32}$$

where $C^{in}(k)$, $C^{out}(k)$, $R^{in}(k)$, $R^{out}(k) \in \mathbb{S}_+^{2n}$, and $\mathcal{E}_{2n}(S)$ is defined as the following $2n$ dimensional ellipsoid

$$\mathcal{E}_{2n}(S) \triangleq \{\tilde{x}^c \in \mathbb{R}^{2n} : (\tilde{x}^c)^T S \tilde{x}^c \leq 1\}. \tag{33}$$

To compute $\mathcal{C}_k$ and $\mathcal{R}_k$, we focus on the ellipsoidal inner and outer approximations of set intersection and the operators Pre and Rch, which are provided by the following proposition and theorem:

Proposition 2. *Let $S_1, S_2 \in \mathbb{R}^{2n}$ be positive semidefinite, then the following set inclusions hold:*

$$\mathcal{E}_{2n}(S_1 + S_2) \subseteq \mathcal{E}_{2n}(S_1) \bigcap \mathcal{E}_{2n}(S_2) \subseteq \mathcal{E}_{2n}(S_1/2 + S_2/2). \tag{34}$$

Theorem 3. *Let $S \in \mathbb{R}^{2n \times 2n}$ be a positive semidefinite matrix. Then the following set inclusions hold:*

$$\mathcal{E}_{2n}(S_p^{in}) \subseteq \mathrm{Pre}(\mathcal{E}_{2n}(S)) \subseteq \mathcal{E}_{2n}(S_p^{out}), \tag{35}$$
$$\mathcal{E}_{2n}(S_r^{in}) \subseteq \mathrm{Rch}(\mathcal{E}_{2n}(S)) \subseteq \mathcal{E}_{2n}(S_r^{out}), \tag{36}$$

where

$$S_p^{in} = f(S),\ S_p^{out} = f(S)/2, \tag{37}$$
$$S_p^{in} = h(S),\ S_p^{out} = h(S)/2, \tag{38}$$

and $f(S), h(S)$ are defined as the following Riccati equations:

$$\begin{aligned} f(S) &\triangleq \tilde{A}^T S \tilde{A} + \tilde{C}^T \tilde{C} \\ &\quad - (\tilde{A}^T S \tilde{B} + \tilde{C}^T \tilde{D})(\tilde{B}^T S \tilde{B} + \tilde{D}^T \tilde{D})^+ (\tilde{B}^T S \tilde{A} + \tilde{D}^T \tilde{C}), \\ h(S) &\triangleq \hat{A}^T S \hat{A} + \hat{C}^T \hat{C} \\ &\quad - (\hat{A}^T S \hat{B} + \hat{C}^T \hat{D})(\hat{B}^T S \hat{B} + \hat{D}^T \hat{D})^+ (\hat{B}^T S \hat{A} + \hat{D}^T \hat{C}). \end{aligned}$$

The matrices $\hat{A} \in \mathbb{R}^{2n \times 2n}$, $\hat{B} \in \mathbb{R}^{2n \times (q+l+2n)}$, $\hat{C} \in \mathbb{R}^{m \times 2n}$, $\hat{D} \in \mathbb{R}^{m \times (q+l+2n)}$ are defined as

$$\begin{aligned} \hat{A} &\triangleq \tilde{A}^+, \qquad \hat{B} \triangleq \left[\ -\tilde{A}^+\tilde{B},\quad I_{2n} - \tilde{A}^+\tilde{A}\ \right], \\ \hat{C} &\triangleq \tilde{C}\tilde{A}^+, \qquad \hat{D} \triangleq \left[\ \tilde{D} - \tilde{C}\tilde{A}^+\tilde{B},\quad \tilde{C} - \tilde{C}\tilde{A}^+\tilde{A}\ \right]. \end{aligned} \tag{39}$$

Proof of Theorem 3. We first prove (37). Consider the augmented set $\mathcal{S}_a \subseteq \mathbb{R}^{2n+q+l}$ of both the state $\tilde{x}^c$ and attacker's action ζ:

$$\mathcal{S}_a = \left\{ \begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix} : \tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{E}_{2n}(S),\ \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \leq 1 \right\}.$$

It is easy to see that $\tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{E}_{2n}(S)$ is equivalent to the following inequality:

$$\begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix}^T \begin{bmatrix} \tilde{A}^T S \tilde{A} & \tilde{A}^T S \tilde{B} \\ \tilde{B}^T S \tilde{A} & \tilde{B}^T S \tilde{B} \end{bmatrix} \begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix} \leq 1.$$

Moreover, $\|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \leq 1$ is equivalent to

$$\begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix}^T \begin{bmatrix} \tilde{C}^T \tilde{C} & \tilde{C}^T \tilde{D} \\ \tilde{D}^T \tilde{C} & \tilde{D}^T \tilde{D} \end{bmatrix} \begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix} \leq 1.$$

Therefore, the augmented set $\mathcal{S}_a$ is the intersection of the following two $2n + q + l$ dimension ellipsoids:

$$\begin{aligned} \mathcal{S}_a = \mathcal{E}_{2n+q+l}\left(\begin{bmatrix} \tilde{A}^T S \tilde{A} & \tilde{A}^T S \tilde{B} \\ \tilde{B}^T S \tilde{A} & \tilde{B}^T S \tilde{B} \end{bmatrix} \right) \\ \bigcap \mathcal{E}_{2n+q+l}\left(\begin{bmatrix} \tilde{C}^T \tilde{C} & \tilde{C}^T \tilde{D} \\ \tilde{D}^T \tilde{C} & \tilde{D}^T \tilde{D} \end{bmatrix} \right). \end{aligned} \tag{40}$$

Thus, by Proposition 2,

$$\mathcal{E}_{2n+q+l}(S_a^{in}) \subseteq \mathcal{S}_a \subseteq \mathcal{E}_{2n+q+l}(S_a^{out}),$$

where

$$\begin{aligned} S_a^{in} &= \begin{bmatrix} \tilde{A}^T S \tilde{A} & \tilde{A}^T S \tilde{B} \\ \tilde{B}^T S \tilde{A} & \tilde{B}^T S \tilde{B} \end{bmatrix} + \begin{bmatrix} \tilde{C}^T \tilde{C} & \tilde{C}^T \tilde{D} \\ \tilde{D}^T \tilde{C} & \tilde{D}^T \tilde{D} \end{bmatrix}, \\ S_a^{out} &= S_a^{in}/2. \end{aligned}$$

Using the Schur complement, we can project the two high dimensional ellipsoids from $\mathbb{R}^{2n+q+l}$ back to $\mathbb{R}^{2n}$ to obtain (37).

We now prove (38). From the definition of Rch, for any $\tilde{x}^c \in \mathrm{Rch}(\mathcal{E}_{2n}(S))$, there exist $\tilde{x}^- \in \mathcal{E}_{2n}(S)$ and ζ, such that

$$\tilde{x}^c = \tilde{A}\tilde{x}^- + \tilde{B}\zeta, \tag{41}$$
$$\|\tilde{C}\tilde{x}^- + \tilde{D}\zeta\| \leq 1. \tag{42}$$

By the properties of the pseudoinverse, we know that $I_{2n} - \tilde{A}^+\tilde{A}$ is a projection from $\mathbb{R}^{2n}$ onto the kernel of A. Thus, (41) can be written as

$$\tilde{x}^- = \tilde{A}^+\tilde{x}^c - \tilde{A}^+\tilde{B}\zeta + (I_{2n} - \tilde{A}^+\tilde{A})\tilde{x}_0,$$

where $\tilde{x}_0 \in \mathbb{R}^{2n}$ is an arbitrary vector. Since $\tilde{x}^- \in \mathcal{E}_{2n}(S)$, we know that

$$\tilde{A}^+\tilde{x}^c - \tilde{A}^+\tilde{B}\zeta + (I_{2n} - \tilde{A}^+\tilde{A})\tilde{x}_0 \in \mathcal{E}_{2n}(S). \tag{43}$$

Furthermore, (42) can be written as

$$\left\|\tilde{C}\tilde{A}^{+}\tilde{x}^{c}+(\tilde{D}-\tilde{C}\tilde{A}^{+}\tilde{B})\zeta+(\tilde{C}-\tilde{C}\tilde{A}^{+}\tilde{A})\tilde{x}_0\right\|\leq 1. \quad (44)$$

By the same argument as the proof of Theorem 3, we can obtain (38). □

The monotonicity of the f and h function is proved in the following theorem:

Theorem 4. *For any $X \geq Y \geq 0$, $f(X) \geq f(Y)$, $h(X) \geq h(Y)$.*

Proof. Let

$$X_a = \begin{bmatrix} \tilde{A}^T X \tilde{A} & \tilde{A}^T X \tilde{B} \\ \tilde{B}^T X \tilde{A} & \tilde{B}^T X \tilde{B} \end{bmatrix} + \begin{bmatrix} \tilde{C}^T \tilde{C} & \tilde{C}^T \tilde{D} \\ \tilde{D}^T \tilde{C} & \tilde{D}^T \tilde{D} \end{bmatrix},$$
$$Y_a = \begin{bmatrix} \tilde{A}^T Y \tilde{A} & \tilde{A}^T Y \tilde{B} \\ \tilde{B}^T Y \tilde{A} & \tilde{B}^T Y \tilde{B} \end{bmatrix} + \begin{bmatrix} \tilde{C}^T \tilde{C} & \tilde{C}^T \tilde{D} \\ \tilde{D}^T \tilde{C} & \tilde{D}^T \tilde{D} \end{bmatrix}.$$

Clearly $X_a \geq Y_a$, which implies that $\mathcal{E}_{2n+q+l}(X_a) \subseteq \mathcal{E}_{2n+q+l}(Y_a)$. Define a projection matrix M as

$$M \triangleq \begin{bmatrix} I_{2n} & 0_{2n\times(q+l)} \end{bmatrix} \in \mathbb{R}^{2n\times(2n+q+l)}.$$

which implies that $f(X) \geq f(Y)$. Similarly, one can prove that $h(X) \geq h(Y)$. □

We are now ready to describe a recursive algorithm to compute the ellipsoidal approximations $C^{in}(k)$, $C^{out}(k)$, $R^{in}(k)$, $R^{out}(k)$. By Theorem 3, we know that $C^{in}(k)$, $C^{out}(k)$ can be evaluated recursively as

$$C^{in}(k+1) = f(C^{in}(k)), C^{out}(k+1) = f(C^{out}(k))/2. \quad (45)$$

Since $C^{in}(1) \geq C^{out}(1) \geq C^{in}(0) = C^{out}(0) = 0$, it is easy to prove by induction that $\{C^{in}(k)\}$ and $\{C^{out}(k)\}$ are monotonically increasing and hence the limits for both sequences exist. Let us denote the limits as C^{in}_{∞} and C^{out}_{∞} respectively. Hence, $R^{in}(k)$ and $R^{out}(k)$ can be computed recursively as

$$\begin{aligned} R^{in}(k+1) &= h(R^{in}(k)) + C^{in}_{\infty}, \\ R^{out}(k+1) &= \left[h(R^{out}(k))/2 + C^{out}_{\infty}\right]/2. \end{aligned} \quad (46)$$

Remark 4. *It is worth noticing that other than the ellipsoidal approximation techniques [18], algorithms such as polyhedral approximation [19] can also be adopted to compute $\mathcal{R}$.*

V. Numerical Examples

Consider the following LTI system:

$$x(k+1) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} x(k) + w(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(k). \quad (47)$$

Suppose two sensors are measuring the first state $x_1(k)$ and the second state $x_2(k)$ respectively. Hence

$$y(k) = x(k) + v(k). \quad (48)$$

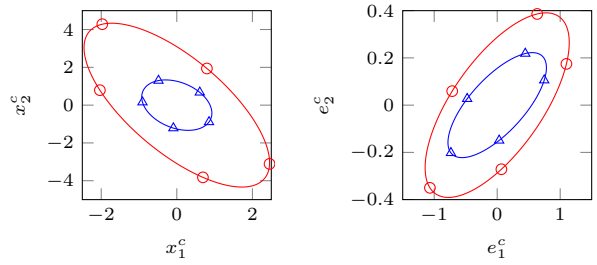
The estimation and control gain matrices are

$$K = \begin{bmatrix} 0.594 & 0.079 \\ 0.079 & 0.694 \end{bmatrix}, L = \begin{bmatrix} -1.244 & -0.422 \end{bmatrix} \quad (49)$$

We consider two cases, where either the first sensor or the second sensor is compromised, i.e. $\Gamma = [1, 0]^T$ or $\Gamma = [0, 1]^T$. We assume that $B^a = 0$ for both cases.

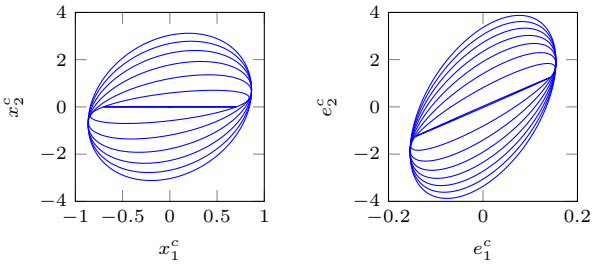
Figure 1 shows the inner and outer approximation of $\mathcal{R}$ when the first sensor is compromised. Since $\mathcal{R}$ is in $\mathbb{R}^4$, we project the ellipsoid to the space of the state $x^c(k)$ and the space of estimation error $e^c(k)$ respectively. From the simulation we can conclude that the reachable region $\mathcal{R}$ is bounded. Therefore the attacker cannot destabilize the system by compromising the first sensor.

Figure 2 shows the inner approximation of $\mathcal{R}_k$ when the second sensor is compromised. It can be seen that $\mathcal{R}_k$ is growing over time. In fact, by Theorem 2 in [11], $\mathcal{R}_k$ must be unbounded. It is worth noticing that by linearity, the reachable set is unbounded for any $\delta > 0$. However, if the system is noiseless, then the adversary need to enforce $\delta = 0$ to avoid detection, which by (18) enforces that $\zeta_k = 0$ for all k. Therefore, no stealthy attack can be launched.



(a) State (b) Estimation Error

Fig. 1. Inner and Outer Approximation of $\mathcal{R}$ When $\Gamma = [1, 0]^T$.



(a) State (b) Estimation Error

Fig. 2. Inner Approximation of $\mathcal{R}_1$ to $\mathcal{R}_7$ When $\Gamma = [0, 1]^T$.

VI. Conclusion

In this paper, we quantify the performance degradation of Cyber-Physical Systems under the effect of stealthy integrity attacks. The CPS is modeled as a stochastic LTI system equipped with a linear filter and feedback controller and χ^2 failure detector. An adversary wishes to induce perturbation in the control loop by compromising a subset of the sensors and injecting an exogenous control input, while remaining stealthy. We show how the attacker's strategy can be formulated as a constrained control problem and that the characterization of the maximum perturbation can be posed as reachable set computation, which can be solved by ellipsoidal approximation methods.

References

[1] T. M. Chen, "Stuxnet, the real start of cyber warfare? [editor's note]," *IEEE Network*, vol. 24, no. 6, pp. 2–3, 2010.

[2] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–6.

[3] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiely, 2009.

[4] K. Zhou, J. C. Doyle, K. Glover *et al.*, *Robust and optimal control*. Prentice Hall New Jersey, 1996, vol. 40.

[5] A. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, Nov 1976.

[6] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, Jan 2010.

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

[8] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. J. Pappas, "The wireless control network: monitoring for malicious behavior," in *IEEE Conference on Decision and Contro*, Atlanta, GA, Dec 2010.

[9] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *Proc. IEEE 51st Annual Conf. Decision and Control (CDC)*, 2012, pp. 3412–3417.

[10] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. 49th IEEE Conf. Decision and Control (CDC)*, 2010, pp. 5967–5972.

[11] Y. Mo and B. Sinopoli, "False data injection attacks in control systems," in *Preprints of the 1st Workshop on Secure Control Systems*, 2010.

[12] J. P. Hespanha, *Linear systems theory*. Princeton university press, 2009.

[13] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," *Automatica*, vol. 7, no. 5, pp. 637–640, Sep. 1971.

[14] P. E. Greenwood and M. S. Nikulin, *A guide to chi-squared testing*. John Wiley & Sons, Apr. 1996.

[15] M. Abramowitz, I. A. Stegun *et al.*, *Handbook of mathematical functions*. Dover New York, 1972, vol. 1, no. 5.

[16] E. Kerrigan, "Robust constraint satisfaction: Invariant sets and predictive control," Ph.D. dissertation, University of Cambridge, 2000.

[17] D. Bertsekas, "Infinite time reachability of state-space regions by using feedback control," *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 604–613, 1972.

[18] D. Angeli, A. Casavola, G. Franz, and E. Mosca, "An ellipsoidal off-line MPC scheme for uncertain polytopic discrete-time systems," *Automatica*, vol. 44, no. 12, pp. 3113 – 3119, 2008.

[19] E. Asarin, O. Bournez, T. Dang, and O. Maler, "Approximate reachability analysis of piecewise-linear dynamical systems," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, N. Lynch and B. Krogh, Eds. Springer Berlin Heidelberg, 2000, vol. 1790, pp. 20–31.

[20] M. A. Goberna, V. Jornet, and M. Rodriguez, "On the characterization of some families of closed convex sets," *Contributions to Algebra and Geometry*, vol. 43, pp. 153–169, 2002.

[21] H. H.Schaefer, *Topological Vector Spaces*. New York: Springer-Verlag, 1971.

APPENDIX A
PROOF OF PROPOSITION 1

This section is devoted to the proof of the last statement in Proposition 1, which requires several intermediate results:

Proposition 3. *Let $\mathcal{S} \subseteq \mathbb{R}^n$ be a closed, convex and symmetric set. Then $\mathcal{S}$ can be decomposed as*

$$\mathcal{S} = \mathcal{K} + \mathcal{V}.$$

where $\mathcal{V}$ is a subspace of $\mathbb{R}^m$ and $\mathcal{K}$ is a compact, convex and symmetric set, which is orthogonal to $\mathcal{V}$.

Proof. The proposition can be proved using Corollary 2.1 in [20]. □

Proposition 4. *The following statements are true:*

1) *Let $\mathcal{K} \subset \mathbb{R}^n$ be a compact (closed and bounded) set and $\mathcal{S}_0 \subseteq \mathbb{R}^n$ be a closed set. Then $\mathcal{S} = \mathcal{K} + \mathcal{S}_0$ is a closed set.*
2) *Let $\mathcal{K} \subset \mathbb{R}^n$ be a compact set and f be a continuous function, then $f(\mathcal{K})$ is also compact.*

Proof. The proof can be found in [21]. □

Lemma 1. *Let $\mathcal{S} \in \mathbb{R}^n$ be a closed, convex and symmetric set, then*

$$\mathrm{Pre}(\mathcal{S}) = \{\tilde{x}^c \in \mathbb{R}^{2n} : \exists \zeta,\ s.t.\ \tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{S}, \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \leq 1\}.$$

is also closed, convex and symmetric.

Proof. One can verify that $\mathrm{Pre}(\mathcal{S})$ is convex and symmetric. Hence, we only need to prove that $\mathrm{Pre}(\mathcal{S})$ is closed. To this end, define:

$$\mathcal{S}_a \triangleq \left\{ \begin{bmatrix} \tilde{x}^c \\ \zeta \end{bmatrix} \in \mathbb{R}^{2n+q+l} : \tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \mathcal{S}, \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \leq 1 \right\}.$$

Since $\tilde{A}\tilde{x}^c + \tilde{B}\zeta$ and $\tilde{C}\tilde{x}^c + \tilde{D}\zeta$ are continuous with respect to $\tilde{x}^c$ and ζ, $\mathcal{S}_a$ is also closed, convex and symmetric. By Proposition 3, we know that $\mathcal{S}_a = \mathcal{K}_a + \mathcal{V}_a$, where $\mathcal{K}_a$ is compact and $\mathcal{V}_a$ is a subspace. Now define a projection matrix M:

$$M \triangleq \begin{bmatrix} I_{2n} & 0_{2n\times(q+l)} \end{bmatrix} \in \mathbb{R}^{2n\times(2n+q+l)}.$$

Thus, $\mathrm{Pre}(\mathcal{S}) = M\mathcal{S}_a = M\mathcal{K}_a + M\mathcal{V}_a$. By Proposition 4, $M\mathcal{K}_a$ is compact. Furthermore, $M\mathcal{V}_a$ is a subspace of $\mathbb{R}^n$ and thus closed. Hence, by Proposition 4, $\mathrm{Pre}(\mathcal{S}) = M\mathcal{K}_a + M\mathcal{V}_a$ is closed. □

We are now ready to prove Proposition 1:

Proof. Since $\mathcal{C}_\infty$ is controlled invariant, $\mathcal{C}_\infty \subseteq \mathrm{Pre}(\mathcal{C}_\infty)$. On the other hand, one can verify that $\mathcal{C}_1 = \mathrm{Pre}(\mathbb{R}^{2n}) \subseteq \mathcal{C}_0$.

Thus, by the monotonicity of Pre, we know that

$$\mathcal{C}_\infty \subseteq \cdots \subseteq \mathcal{C}_1 \subseteq \mathcal{C}_0. \tag{50}$$

Hence, we only need to prove that $\bigcap_{i=0}^\infty \mathcal{C}_i \subseteq \mathcal{C}_\infty$. Let $\tilde{x}^c \in \bigcap_{i=0}^\infty \mathcal{C}_i$. From definition, there exist ζ_i, $i \in \mathbb{N}$, such that

$$\tilde{A}\tilde{x}^c + \tilde{B}\zeta_i \in \mathcal{C}_{i-1},\ \|\tilde{C}\tilde{x}^c + \tilde{D}\zeta_i\| \leq 1.$$

Such ζ_is may not be unique. As a result, we will choose those ζ_is with minimum norm. By Lemma 1 and 3, we know that $\mathcal{C}_i$ can be written as $\mathcal{C}_i = \mathcal{K}_i + \mathcal{V}_i$, where $\mathcal{K}_i$ is compact and $\mathcal{V}_i$ is a subspace. Now by (50),

$$\mathcal{V}_0 \supseteq \mathcal{V}_1 \supseteq \mathcal{V}_2 \supseteq \ldots$$

Let us define subspace

$$\mathcal{V} \triangleq \bigcap_{i=0}^\infty \mathcal{V}_i.$$

Since $\mathcal{V}_i$ is of finite dimension, there must exists an N, such that $\mathcal{V}_i = \mathcal{V}$ for all $i \geq N$, which further implies that $\mathcal{K}_i \supseteq \mathcal{K}_{i+1}$, $i \geq N$. Hence, $\mathcal{K}_i$ is uniformly bounded.

Now we want to prove that $\|\zeta_i\|$ is bounded. Consider the opposite. By Bolzano-Weierstrass Theorem, there exists a subsequence $\{\zeta_{i_j}\}$, such that

$$\lim_{j\to\infty} \|\zeta_{i_j}\| = \infty,\ \lim_{j\to\infty} \zeta_{i_j}/\|\zeta_{i_j}\| = v.$$

Hence

$$\tilde{B}\frac{\zeta_{i_j}}{\|\zeta_{i_j}\|} \in \mathcal{V}_i + \frac{1}{\|\zeta_{i_j}\|}(\mathcal{K}_i - \tilde{A}\tilde{x}^c),\ \left\|\tilde{D}\frac{\zeta_{i_j}}{\|\zeta_{i_j}\|}\right\| \leq \frac{1}{\|\zeta_{i_j}\|}(1 + \|\tilde{C}\tilde{x}^c\|),$$

which implies that $\tilde{B}v \in \mathcal{V}, \tilde{D}v = 0$. Therefore, for any $\alpha \in \mathbb{R}$,

$$\tilde{A}\tilde{x}^c + \tilde{B}(\zeta_i + \alpha v) \in \mathcal{C}_{i-1},\ \|\tilde{C}\tilde{x}^c + \tilde{D}(\zeta_i + \alpha v)\| \leq 1.$$

As a result, the fact that $\|\zeta_i\|$ is unbounded contradicts with the minimality of $\|\zeta_i\|$. Thus, $\|\zeta_i\|$ must be bounded. Now by Bolzano-Weierstrass Theorem, there exists a subsequence $\{\zeta_{i_j}\}$, such that

$$\lim_{j\to\infty} \zeta_{i_j} = \zeta.$$

It is easy to see that $\|\tilde{C}\tilde{x}^c + \tilde{D}\zeta\| \leq 1$. On the other hand, for any $j > i$, $\tilde{A}\tilde{x}^c + \tilde{B}\zeta_j \in \mathcal{C}_{j-1} \subseteq \mathcal{C}_i$. Since $\mathcal{C}_i$ is closed, we know that

$$\tilde{A}\tilde{x}^c + \tilde{B}\zeta \in \bigcap_{i=0}^\infty \mathcal{C}_i.$$

Thus, $\bigcap_{i=0}^\infty \mathcal{C}_i$ is controlled invariant. Since $\mathcal{C}_\infty$ is the largest controlled invariant set, $\bigcap_{i=0}^\infty \mathcal{C}_i \subseteq \mathcal{C}_\infty$, which concludes the proof. □