Stealthy Adversaries Against Uncertain Cyber-Physical Systems: Threat of Robust Zero-Dynamics Attack

Gyunghoon Park, Chanhwa Lee, Hyungbo Shim, *Senior Member, IEEE*, Yongsoon Eun, *Member, IEEE*, and Karl H. Johansson, *Fellow, IEEE*

***Abstract*—In this paper, we address the problem of constructing a *robust* stealthy attack that compromises *uncertain* cyber-physical systems having unstable zeros. We first interpret the (non-robust) conventional zero-dynamics attack based on Byrnes–Isidori normal form, and then present a new *robust zero-dynamics attack* for uncertain plants. Different from the conventional strategy, our key idea is to isolate the real zero-dynamics from the plant's input–output relation and to replace it with an auxiliary nominal zero-dynamics. As a result, this alternative attack does not require the exact model knowledge anymore. The price to pay for the robustness is to utilize the input and output signals of the system (i.e., disclosure resources). It is shown that a disturbance observer can be employed to realize the new attack philosophy when there is a lack of model knowledge. Simulation results with a hydro-turbine power system are presented to verify the attack performance and robustness.**

***Index Terms*—Disturbance observer, robustness, security, uncertain system, zero-dynamics attack.**

Manuscript received May 3, 2018; revised December 2, 2018; accepted February 3, 2019. Date of publication March 7, 2019; date of current version December 3, 2019. This work was supported in part by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2014-0-00065, Resilient Cyber-Physical Systems Research), and in part by National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. NRF-2017R1E1A1A03070342). K. H. Johansson was partially supported by the Knut and Alice Wallenberg Foundation, the Swedish Strategic Research Foundation, and the Swedish Research Council. Recommended by Associate Editor Z. Wang. This paper was presented in part at the material in this paper was partially presented at the 55th IEEE Conference on Decision and Control, Las Vegas, USA, December 2016 [1]. *(Corresponding author: Hyungbo Shim.)*

G. Park is with the Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul 136-791, Korea (e-mail: gyunghoon.p@gmail.com).

C. Lee is with the Research & Development Division, Hyundai Motor Company, Gyeonggi-do 18280, Korea (e-mail: chanhwa.lee@gmail.com).

H. Shim is with the ASRI, Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea (e-mail: hshim@snu.ac.kr).

Y. Eun is with the Department of Information & Communication Engineering, Daegu Gyeongbuk Institute of Science & Technology, Daegu 47988, Korea (e-mail: yeun@dgist.ac.kr).

K. H. Johansson is with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, 144 28 Stockholm, Sweden (e-mail: kallej@kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2019.2903429

I. Introduction

Modern control systems often have complex structures integrating physical plants and digital devices, which are linked through communication networks. These cyber-physical systems (CPS) offer great opportunities to achieve high cost efficiency and productivity over traditional control systems [2]–[4].

Yet at the same time, CPS are vulnerable to malicious attackers, as nowadays the data networks are easier to access by anonymous users. Serious cyber threats to CPS already have taken place in recent years, such as the attacks on the U.S. electric grid [5] and the Stuxnet malware [6]. In this context, it is not surprising that security of the CPS has attracted widespread attention with emerging resilient control and secure estimation schemes [7]–[12].

With the increased interest, a variety of cyber attack scenarios have been studied from a control-theoretic perspective; e.g., denial-of-service (DoS) attack, replay attack [13], zero-dynamics attack [7], [8], [14], bias injection attack [7], optimal linear attack [15], switching location attack [16], multi-rate sampling attack [17], to name just a few. Among various purposes of these attacks, *stealthiness* is of utter importance to most adversaries: i.e., when an attack signal enters CPS, its impact should not be detected by any anomaly detector. In view of the adversary, one approach for achieving stealthiness is to employ structural information of the plant in the attack design. For instance, zero-dynamics attack is known as a model-based attack strategy that remains stealthy [7], [8], [14]. In this attack scenario, the adversary duplicates exactly the real unstable zero-dynamics of non-minimum phase plants. As a result, the attack signal conceals itself in the so-called output-nulling space, even if a large amount of false data are injected into the plant [7], [14].

Model-based attacks may easily lose their stealthiness when the model knowledge is not perfect. Indeed, even small mismatch between the real and estimated models leaves the zero-dynamics attack detectable [14]. This fundamental limitation of model-based attacks has led to recent developments of attack prevention strategies, such as structural modification schemes [14], [18]. Furthermore, exact model knowledge is not obtainable in many industrial problems, which is another hurdle to the attacker. If so, are CPS be safe from these lethal stealthy attacks thanks to model uncertainty?

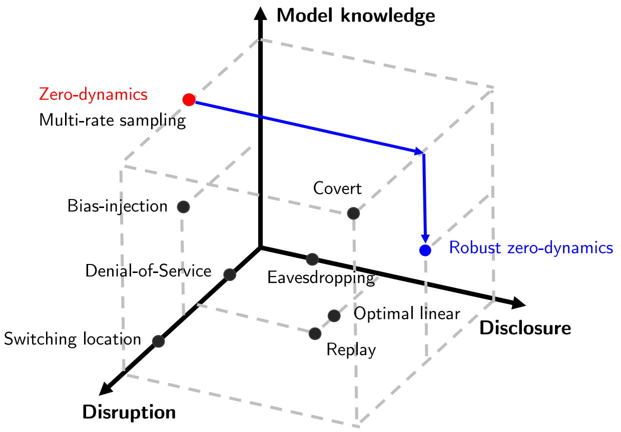


Fig. 1. Cyber-physical attack space [10] with model knowledge, disruption, and disclosure resources: The robust zero-dynamics attack is at entirely new location.

Interestingly, we find in this paper that it may not be the case when the attacker employs robust control techniques in their attack designs. Specifically, we address the problem of constructing a *robust zero-dynamics attack* that is stealthy for *uncertain* non-minimum phase plants. Moving away from the traditional methods, our key idea is: 1) to eliminate the effect of model uncertainty and the real zero-dynamics from the input–output relation in the plant's dynamics, and 2) to build up an *auxiliary* nominal zero-dynamics which replaces the role of the real counterpart. Then the actual zero-dynamics is left alone while being unstable, and thus its state trajectory will diverge without being detected. This new philosophy is realized by regarding the terms to be eliminated as a so-called *lumped disturbance*, and by designing a robust controller that estimates and compensates the lumped disturbance. Actually, all this can be done by disturbance observer [19], [20], which will play the role of attack generator in this work. It is worth mentioning that the price to pay for less model knowledge is the necessity of the control input and the plant's output information; in other words, the proposed robust zero-dynamics attack requires more *disclosure resources* [10], as depicted in Fig. 1.

Notation: For column vectors a and b, we write $[a; b]$ for $\mathrm{col}(a, b) = [a^\top, b^\top]^\top$. For two sets $\mathcal{A}$ and $\mathcal{B}$, the distance between the sets is defined as $\mathrm{dist}(\mathcal{A}, \mathcal{B}) := \inf_{a\in\mathcal{A}, b\in\mathcal{B}} \|a - b\|$. In addition, $\mathcal{A}$ is said to be strictly larger than $\mathcal{B}$ if $\mathcal{A} \supset \mathcal{B}$ and $\mathrm{dist}(\partial\mathcal{A}, \partial\mathcal{B}) > 0$ where $\partial\mathcal{A}$ indicates the boundary of $\mathcal{A}$. For a square matrix M, $\Lambda(M)$ indicates the set of the eigenvalues of M. The zero vector is denoted by $0_k \in \mathbb{R}^k$, while $0_{k\times m} \in \mathbb{R}^{k\times m}$ and $I_k \in \mathbb{R}^{k\times k}$ indicate the zero and identity matrices, respectively. For simplicity, we often write them without the subscripts if their dimensions are obvious.

II. Normal Form-Based Interpretation of Zero-Dynamics Attack

Zero-dynamics attack is a systematic methodology to compromise a class of CPS whose physical plants are of non-minimum phase [7], [9], [14], including a variety of modern control systems such as power generating systems [22], autopilots for unmanned automated vehicles [23], water level control systems [24], [25], to name just a few. The basic concept of the attack is that the attack generator produces a signal based on the unstable zero-dynamics of the physical plant and injects its diverging output into the actuator channel. This consequently leads to two important features: 1) the actual (zero-dynamics) state diverges as time elapses, and 2) the plant's state remains close to the output-nulling space so that the corresponding output is almost zero. By the latter property, the zero-dynamics attack has been known as a *stealthy attack*.

In this section, we re-interpret the zero-dynamics attack. In particular, while the geometric control theory [7] usually has been employed as a tool for the analysis in the literature, we here present another way to analyze the attack, based on the *Byrnes–Isidori normal form* [21, Ch. 13]. This new approach will allow us to gain further insight on the attack, especially on its fundamental limitation against model uncertainty.

A. System Description

Consider a linear single-input single-output (SISO) plant under an actuator attack, especially represented in the Byrnes–Isidori normal form[1]

$$\dot{z} = Sz + Gy, \tag{1a}$$

$$\dot{x} = \mathsf{A}_\nu x + \mathsf{B}_\nu\left(\psi^\top z + \phi^\top x + g(u + a)\right), \tag{1b}$$

$$y = \mathsf{C}_\nu x$$

where $z \in \mathbb{R}^\mu$ and $x \in \mathbb{R}^{n-\mu}$ are the states with the relative degree $\nu := n - \mu \geq 1$, $y \in \mathbb{R}$ is the output, $u \in \mathbb{R}$ is the control input, and $a \in \mathbb{R}$ is the attack signal that enters the actuator channel. For an integer $i \geq 1$, the matrices A_i, B_i, and C_i are given by

$$\mathsf{A}_i := \begin{bmatrix} 0_{i-1} & I_{i-1} \\ 0 & 0_{i-1}^\top \end{bmatrix}, \quad \mathsf{B}_i := \begin{bmatrix} 0_{i-1} \\ 1 \end{bmatrix}, \quad \mathsf{C}_i := \begin{bmatrix} 1 & 0_{i-1}^\top \end{bmatrix}.$$

The matrices S, G, ψ, and ϕ, and the scalar g are with suitable dimensions. Without loss of generality, it is supposed that (1) is controllable and observable, and the high-frequency gain g is positive.

For now, it is assumed that at least one of the eigenvalues of S lies in the open right half-plane (so that the plant (1) is of non-minimum phase). Then without loss of generality, the z-dynamics (1a) can be rewritten (by applying a suitable coordinate change for z) as

$$\begin{bmatrix} \dot{z}_\mathsf{u} \\ \dot{z}_\mathsf{s} \end{bmatrix} = \begin{bmatrix} S_\mathsf{u} & 0 \\ 0 & S_\mathsf{s} \end{bmatrix} \begin{bmatrix} z_\mathsf{u} \\ z_\mathsf{s} \end{bmatrix} + \begin{bmatrix} G_\mathsf{u} \\ G_\mathsf{s} \end{bmatrix} y \tag{2}$$

where $S_\mathsf{u} \in \mathbb{R}^{\mu_\mathsf{u}\times\mu_\mathsf{u}}$ and $S_\mathsf{s} \in \mathbb{R}^{\mu_\mathsf{s}\times\mu_\mathsf{s}}$ are square matrices with $\mu_\mathsf{u} \geq 1$ and $\mu_\mathsf{s} := \mu - \mu_\mathsf{u}$, such that all the eigenvalues of S_u and S_s are located in the open right half-plane and the closed left

[1] Any (strictly proper) SISO linear system can be transformed into the Byrnes–Isidori normal form [21, Ch. 13]. In this form, the zeros of the transfer function of the SISO linear system coincide with the eigenvalues of S. For this reason, $\dot{z} = Sz$ is called the *zero-dynamics*.

half-plane, respectively. The matrices G_{u} and G_{s} are constant and satisfy $G = [G_{\mathsf{u}}; G_{\mathsf{s}}]$.

The control input u in (1) is supposed to be generated by an output feedback controller

$$\dot{c} = Pc + Q(r - y), \quad u = Jc + K(r - y). \tag{3}$$

Here $c \in \mathbb{R}^m$ is the controller state, and $r \in \mathbb{R}$ is the reference signal, which is bounded and sufficiently smooth with bounded time derivatives, and P, Q, J, and K are some constant matrices. We assume that without the attack (i.e., $a(t) \equiv 0$), the closed-loop system (1) and (3) is stable. Note that we are not assuming the plant (1) is stable.

As introduced in the previous works [7], [14], the zero-dynamics attack is usually constructed by duplicating the zero-dynamics of the plant (1). In particular, with the help of the normal form representation, one can express the attack as

$$\dot{z}^{\mathsf{a}} = Sz^{\mathsf{a}}, \quad a_{\mathsf{za}} = -\frac{1}{g}\psi^\top z^{\mathsf{a}} \tag{4}$$

where $z^{\mathsf{a}} =: [z_{\mathsf{u}}^{\mathsf{a}}; z_{\mathsf{s}}^{\mathsf{a}}] \in \mathbb{R}^{\mu_{\mathsf{u}}+\mu_{\mathsf{s}}}$ is the attacker's state. (In what follows, the superscript "a" is used to indicate signals generated by the adversary.) To activate the unstable mode of the z^{a}-dynamics (4), the initial condition $z_{\mathsf{u}}^{\mathsf{a}}(t_0)$ of the unstable part is selected as a nonzero vector. (Hereinafter, we denote the moment when the attack $a(t)$ enters the system as $t = t_0$.)

B. Performance of Zero-Dynamics Attack

We start the analysis of the attack (4) with a new variable $\chi := [x; c] \in \mathbb{R}^{\nu+m}$. Then the actual stable closed-loop system (1) and (3) can be rewritten in a more compact form

$$\dot{z} = Sz + GC\chi, \tag{5a}$$

$$y = C\chi, \quad \dot{\chi} = A\chi + Er + B\left(ga + \psi^\top z\right) \tag{5b}$$

where the matrices A, B, C, and E are given by

$$A := \begin{bmatrix} \mathsf{A}_\nu + \mathsf{B}_\nu\left(\phi^\top - gK\mathsf{C}_\nu\right) & g\mathsf{B}_\nu J \\ -Q\mathsf{C}_\nu & P \end{bmatrix}, \tag{6a}$$

$$B := \begin{bmatrix} \mathsf{B}_\nu \\ 0_m \end{bmatrix}, \quad E := \begin{bmatrix} g\mathsf{B}_\nu K \\ Q \end{bmatrix}, \quad C := \begin{bmatrix} \mathsf{C}_\nu & 0_m^\top \end{bmatrix}. \tag{6b}$$

For comparison, by putting $a(t) \equiv 0$ into (5) we obtain an (auxiliary) *attack-free* closed-loop system

$$\dot{z}_{\mathsf{o}} = Sz_{\mathsf{o}} + GC\chi_{\mathsf{o}}, \tag{7a}$$

$$y_{\mathsf{o}} = C\chi_{\mathsf{o}}, \quad \dot{\chi}_{\mathsf{o}} = A\chi_{\mathsf{o}} + Er + B\psi^\top z_{\mathsf{o}} \tag{7b}$$

where $\chi_{\mathsf{o}} =: [x_{\mathsf{o}}; c_{\mathsf{o}}]$ is the attack-free counterpart of χ.

Now, the nature of the zero-dynamics attack (4) is introduced in the following proposition.

Proposition 1: The solution $[z^{\mathsf{a}}(t); z(t); \chi(t)]$ of the closed-loop system (5) under the attack $a = a_{\mathsf{za}}$ in (4), initiated in $\mathbb{R}^\mu \times \mathbb{R}^\mu \times \mathbb{R}^{\nu+m}$, satisfies the following statements:

a) For any $z^{\mathsf{a}}(t_0) \in \mathbb{R}^\mu$ such that $z_{\mathsf{u}}^{\mathsf{a}}(t_0) \neq 0$,

$$\|z_{\mathsf{u}}(t)\| \to \infty \text{ as } t \to \infty; \tag{8}$$

b) For the solution $[z_{\mathsf{o}}; \chi_{\mathsf{o}}]$ of the attack-free system (7) initiated at $[z_{\mathsf{o}}(t_0); \chi_{\mathsf{o}}(t_0)] = [z(t_0); \chi(t_0)]$,

$$\left\|\chi(t) - \chi_{\mathsf{o}}(t)\right\| \leq k_1 e^{-\lambda_1(t-t_0)}\|z^{\mathsf{a}}(t_0)\|, \quad \forall t \geq t_0$$

where $k_1 > 0$ and $\lambda_1 > 0$ are constant. ■

Proposition 1 explicitly points out that the zero-dynamics attack (4) is capable of damaging the internal state $z_{\mathsf{u}}(t)$ of the plant. We also note that just a small non-zero initial condition $\|z_{\mathsf{u}}^{\mathsf{a}}(t_0)\|$ of the attack generator (4) will do the job, while maintaining stealthiness in the sense that

$$\|y(t) - y_{\mathsf{o}}(t)\| = \|C\chi(t) - C\chi_{\mathsf{o}}(t)\| < \epsilon, \quad \forall t \geq t_0 \tag{9}$$

with a given threshold $\epsilon > 0$.

Proof: Let us define an error variable $\tilde{z} := z - z^{\mathsf{a}}$. It is trivial that $ga + \psi^\top z = \psi^\top \tilde{z}$. It then follows that the closed-loop system (5) is transformed into

$$\dot{\tilde{z}} = \dot{z} - \dot{z}^{\mathsf{a}} = S(z - z^{\mathsf{a}}) + GC\chi = S\tilde{z} + GC\chi, \tag{10a}$$

$$\dot{\chi} = A\chi + Er + B\psi^\top \tilde{z} \tag{10b}$$

which has exactly the same (stable) dynamics as the attack-free one (7). Thus for some positive constants k_1 and λ_1,

$$\begin{aligned} &\left\|[\tilde{z}(t); \chi(t)] - [z_{\mathsf{o}}(t); \chi_{\mathsf{o}}(t)]\right\| \\ &\leq k_1 e^{-\lambda_1(t-t_0)}\left\|[\tilde{z}(t_0); \chi(t_0)] - [z_{\mathsf{o}}(t_0); \chi_{\mathsf{o}}(t_0)]\right\| \\ &= k_1 e^{-\lambda_1(t-t_0)}\|z^{\mathsf{a}}(t_0)\| \end{aligned}$$

where the last equality results from $z(t_0) = z_{\mathsf{o}}(t_0)$ and $\chi(t_0) = \chi_{\mathsf{o}}(t_0)$. This directly implies the item (b). On the other hand, the attacker's state $z^{\mathsf{a}}(t)$ generated by (4) with nonzero $z_{\mathsf{u}}^{\mathsf{a}}(t_0)$ must diverge as time goes on. It means that $z(t) = z^{\mathsf{a}}(t) + \tilde{z}(t)$ also diverges, because the state $\tilde{z}(t)$ of the stable system (10) remains bounded. This completes the proof. ■

Remark 1: From the analysis, it is clear that the lower-order dynamics $\dot{z}_{\mathsf{u}}^{\mathsf{a}} = S_{\mathsf{u}} z_{\mathsf{u}}^{\mathsf{a}}$ and $a_{\mathsf{za}} = -(1/g)\psi_{\mathsf{u}}^\top z_{\mathsf{u}}^{\mathsf{a}}$ (where ψ_{u} is a suitable partition of ψ) is enough to realize the zero-dynamics attack. ■

C. Limitation of Zero-Dynamics Attack Against Model Uncertainty

It is important to note that exact model knowledge on the plant (1) is of necessity in the design of the zero-dynamics attack (4). However, such a requirement is quite unrealistic. This is because in most industrial systems, it is not always possible for the attacker (nor for the defender) to obtain the exact plant model. In other words, it is natural to assume that the physical plant (1) has model uncertainty.

Assumption 1: The parameters S, G, ψ, ϕ, and g of the plant (1) are uncertain, while the uncertain quantities are bounded and their bounds are known to the attacker.[2] In particular, $0 < \underline{g} \leq g \leq \overline{g}$ for some constants $\underline{g}$ and $\overline{g}$. ■

We assume that the controller (3) is appropriately designed to robustly stabilize the uncertain plant satisfying Assumption 1.

[2] The attacker need not know exact bounds of uncertainties. Overestimate would work.

From now on, we take a glance at the situation when the attacker tries to design a zero-dynamics attack, in the presence of the model uncertainty in Assumption 1. Since the ideal structure (4) is not available at this stage, a possible alternative would be

$$\dot{z}^{\mathsf{a}} = S_{\mathsf{n}} z^{\mathsf{a}}, \quad a_{\mathsf{za}} = -\frac{1}{g_{\mathsf{n}}} \psi_{\mathsf{n}}^{\top} z^{\mathsf{a}} \tag{11}$$

where S_{n}, ψ_{n}, and $g_{\mathsf{n}} > 0$ are selected as some nominal counterparts of S, ψ, and g, respectively. Then, the attack signal $a_{\mathsf{za}}(t)$ is exponentially diverging with rate determined by the unstable modes of S_{n}. If these modes are different from the zeros of (5) (i.e., the eigenvalues of S), then the output $y(t)$ should diverge with the same exponential rate as those modes. In fact, even arbitrarily small differences will lead to a diverging output with fixed rate.

From the discussion so far, it may seem that CPS are safe from those stealthy attacks because model uncertainty exists in practice. Unfortunately, we find in the next section that there is another type of stealthy attack which is *robust against model uncertainty*.

Remark 2: When the system (1) has a non-trivial transfer function and is controllable but unobservable, then the unobservable eigenvalues are the transmission zeros. If those zeros are unstable, then almost all signals $a(\cdot)$ can drive the unobservable states unbounded. An example is when $\phi = 0$ and S is not Hurwitz in (1). This attack does not require the model knowledge, but we exclude these cases by assuming both controllability and observability because such systems cannot be stabilized by feedback. ■

III. Robust Zero-Dynamics Attack for Uncertain Cyber-Physical Systems

A. Problem Formulation

In what follows, we consider the closed-loop system (1) and (3) (or equivalently, (5)) where the plant (1) of interest has parametric uncertainty as in Assumption 1. Also, we suppose that the initial conditions $z(t_0)$ and $\chi(t_0)$ of the closed-loop system (5) belong to some compact sets $\mathcal{Z}_0 \subset \mathbb{R}^{\mu}$ and $\mathcal{X}_0 \subset \mathbb{R}^{\nu+m}$, respectively.

The following (attack-free) *nominal* plant of (1) is taken into account:

$$\begin{aligned} \dot{z}_{\mathsf{n}} &= S_{\mathsf{n}} z_{\mathsf{n}} + G_{\mathsf{n}} y_{\mathsf{n}}, \\ \dot{x}_{\mathsf{n}} &= \mathsf{A}_{\nu} x_{\mathsf{n}} + \mathsf{B}_{\nu} \left(\psi_{\mathsf{n}}^{\top} z_{\mathsf{n}} + \phi_{\mathsf{n}}^{\top} x_{\mathsf{n}} + g_{\mathsf{n}} u_{\mathsf{n}} \right), \\ y_{\mathsf{n}} &= \mathsf{C}_{\nu} x_{\mathsf{n}} \end{aligned} \tag{12}$$

where $z_{\mathsf{n}} \in \mathbb{R}^{\mu}$ and $x_{\mathsf{n}} \in \mathbb{R}^{\nu}$ are the nominal states, $y_{\mathsf{n}} \in \mathbb{R}$ is the nominal output, and $u_{\mathsf{n}} \in \mathbb{R}$ is the nominal input, which is generated by the existing control law (3) as

$$\dot{c}_{\mathsf{n}} = P c_{\mathsf{n}} + Q(r - y_{\mathsf{n}}), \quad u_{\mathsf{n}} = J c_{\mathsf{n}} + K(r - y_{\mathsf{n}}) \tag{13}$$

where c_{n} represents the nominal state of the controller. The parameters S_{n}, G_{n}, ψ_{n}, ϕ_{n}, and $g_{\mathsf{n}} > 0$ are nominal counterparts of the actual (uncertain) S, G, ψ, ϕ, and $g > 0$, respectively, and these are the attacker's selection. The nominal model (12) is of course different from the real plant (1), but it is assumed that the parameters of (12) are within the uncertainty bounds of Assumption 1, so that both (1) and (12) have the same relative degree and the same sign of high-frequency gains g and g_{n}. It will be seen that the plant (1) behaves like the nominal model (12) by the initiation of the proposed attack. In this context, it may be better for stealthiness if the nominal model (12) coincides with a design model used for designing the controller (3) (which, however, requires that the design model is leaked to the attacker *a priori*). For brevity, we often express the nominal closed-loop system (12) and (13) with $\chi_{\mathsf{n}} := [x_{\mathsf{n}}; c_{\mathsf{n}}]$ as

$$\dot{z}_{\mathsf{n}} = S_{\mathsf{n}} z_{\mathsf{n}} + G_{\mathsf{n}} C \chi_{\mathsf{n}}, \tag{14a}$$

$$y_{\mathsf{n}} = C \chi_{\mathsf{n}}, \quad \dot{\chi}_{\mathsf{n}} = A_{\mathsf{n}} \chi_{\mathsf{n}} + E_{\mathsf{n}} r + B \psi_{\mathsf{n}}^{\top} z_{\mathsf{n}} \tag{14b}$$

where A_{n} and E_{n} are the same as A and E defined in (6), with S, G, ψ, ϕ, and g being replaced by their nominal counterparts.

We note in advance that similar to (7), the nominal closed-loop system (14) (or equivalently (12) and (13)) will play the role of a reference system in the attack design. For this, the nominal closed-loop system (14) is supposed to be stable; i.e., the matrix

$$\begin{bmatrix} S_{\mathsf{n}} & G_{\mathsf{n}} C \\ B \psi_{\mathsf{n}}^{\top} & A_{\mathsf{n}} \end{bmatrix} \tag{15}$$

is Hurwitz.

Now, motivated by Proposition 1, we formulate the problem of our interest with respect to the uncertain plant.

Problem Statement: For given $\underline{z}_{\mathsf{u}} > 0$ and $\epsilon > 0$, construct a *robust* attack generator

$$\dot{q}^{\mathsf{a}} = \Phi(q^{\mathsf{a}}, u, y), \quad a = \Psi(q^{\mathsf{a}}, u, y) \tag{16}$$

that achieves the following properties simultaneously for all admissible model uncertainties in Assumption 1:

a) $\|z_{\mathsf{u}}(t)\|$ becomes eventually larger than $\underline{z}_{\mathsf{u}} > 0$ within a finite time $t = t_{\mathsf{fin}} \geq t_0$;

b) $\|y(t) - y_{\mathsf{n}}(t)\|$ is smaller than the threshold $\epsilon > 0$ until the attack succeeds (i.e., for all $t_0 \leq t \leq t_{\mathsf{fin}}$). ■

The items in Problem Statement can be interpreted as some refinements of those in Proposition 1. On one hand, item (a) indicates the capability of the attack (16) to damage the plant's internal state $z(t)$. Here, $\underline{z}_{\mathsf{u}}$ is one of the attacker's design specifications. On the other hand, a *new* notion of stealthiness is introduced in item (b). Indeed, the actual output $y(t)$ under attack is compared with the output $y_{\mathsf{n}}(t)$ of the *nominal* system (14), rather than with $y_{\mathsf{o}}(t)$ of the (attack-free) *uncertain* system (5) as in Proposition 1. At first glance, it may seem that item (b) easily fails if the model uncertainty is large. However, this is often not the case even for large model uncertainty, as long as the existing controller (3) is robust against the parametric uncertainties of Assumption 1. In fact, when a tracking or regulating problem is (robustly) solved by (3) for both actual and nominal systems (with no attack), their outputs $y_{\mathsf{n}}(t)$ and $y_{\mathsf{o}}(t)$ reach the same reference $r(t)$ in the end. It means that $y_{\mathsf{n}}(t) \approx y_{\mathsf{o}}(t)$ during the steady-state operation of the system when the attack is usually initiated. In summary, we claim in this paper that the new stealthiness is also valid if the uncertainty is not large or if the attack (16) enters the system in the steady state.

Remark 3: The stealthiness defined in Problem Statement is basically approximate in the sense that $t_{\mathsf{fin}} \neq \infty$ and $\epsilon \neq 0$. Nonetheless, this approximation is enough for the attackers to remain undetected in practice, as long as: 1) the threshold ϵ is of the order of measurement noise, and 2) the attack detector is designed based on the corrupted output measurement $y(t)$; e.g., the observer-based fault detector in [26] consisting of a Luenberger observer

$$\begin{bmatrix}\dot{\hat{z}}\\ \dot{\hat{x}}\end{bmatrix} = \hat{A}_{\mathsf{n}}\begin{bmatrix}\hat{z}\\ \hat{x}\end{bmatrix} + \hat{B}_{\mathsf{n}}u - \hat{L}_{\mathsf{n}}\left(\hat{C}_{\mathsf{n}}\begin{bmatrix}\hat{z}\\ \hat{x}\end{bmatrix} - y\right) \tag{17a}$$

for the (attack-free) nominal model (12) of (1), and a residual signal

$$\hat{r}_{\mathsf{n}}(t) = \hat{C}_{\mathsf{n}}\begin{bmatrix}\hat{z}(t)\\ \hat{x}(t)\end{bmatrix} - y(t) \tag{17b}$$

to be monitored, in which

$$\hat{A}_{\mathsf{n}} := \begin{bmatrix} S_{\mathsf{n}} & G_{\mathsf{n}}\mathsf{C}_\nu \\ \mathsf{B}_\nu \psi_{\mathsf{n}}^\top & \mathsf{A}_\nu + \mathsf{B}_\nu \phi_{\mathsf{n}}^\top \end{bmatrix},$$

$\hat{B}_{\mathsf{n}} := [0_\nu; g_{\mathsf{n}}\mathsf{B}_\nu]$, $\hat{C}_{\mathsf{n}} := [0_{n-\nu}^\top, \mathsf{C}_\nu]$, and $\hat{L}_{\mathsf{n}}$ is the observer gain matrix satisfying that $\hat{A}_{\mathsf{n}} - \hat{L}_{\mathsf{n}}\hat{C}_{\mathsf{n}}$ is Hurwitz. Indeed, under the item (b) of Problem Statement, one can readily rewrite the attacked output $y(t)$ as $y(t) = y_{\mathsf{n}}(t) + w(t)$ where $w(t)$ is a (noise-like) signal such that $\|w(t)\| < \epsilon$ for all $t_0 \leq t \leq t_{\mathsf{fin}}$. In that case, with the output feedback-based detectors like (17), it is hardly possible to distinguish the effect of the attack from that of the actual noise (at least until the attack succeeds) [26]. Similar conclusion can be made with the conventional zero-dynamics attack in Section II, which satisfies (9). ■

We further remark that, different from traditional zero-dynamics attack (4), the attack generator (16) explicitly makes use of the signals u and y. This is in fact the price to pay for the *robustness* against model uncertainty; i.e., instead of using less model knowledge, the attacker relies more on the input and output information of the plant to adjust to uncertain environment on-line. In short, more disclosure resources are needed as follows.

Assumption 2: The plant output $y(t)$ and the control input $u(t)$ are available to attackers. ■

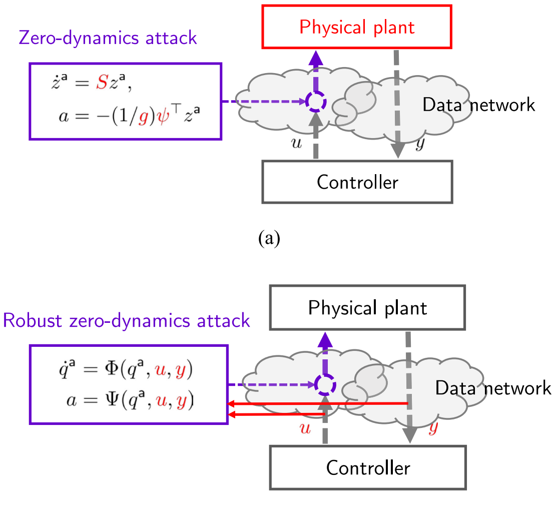
In addition, for a technical reason, we restrict our attention on the non-minimum phase systems with *hyperbolic* zero-dynamics.

Assumption 3: At least one of the eigenvalues of S lies in the open right half-plane, and none of the eigenvalues are located on the imaginary axis of the complex plane. ■

To distinguish (16) from the (non-robust) zero-dynamics attack (4), we call (16) a *robust zero-dynamics attack*. Overall configurations of these attack scenarios are depicted in Fig. 2.

B. A New Attack Policy on Unstable Zero-Dynamics

As an intermediate step, in this subsection we provide a new attack strategy on the non-minimum phase plant (1). It is noted in advance that the method to be provided here is not realizable yet, but we will shortly make it feasible in the next section.


Zero-dynamics attack
$\dot{z}^{\mathsf{a}} = Sz^{\mathsf{a}}$,
$a = -(1/g)\psi^\top z^{\mathsf{a}}$
Physical plant
Data network
u
y
Controller
Robust zero-dynamics attack
$\dot{q}^{\mathsf{a}} = \Phi(q^{\mathsf{a}}, u, y)$
$a = \Psi(q^{\mathsf{a}}, u, y)$
Physical plant
Data network
u
y
Controller

(a)

(b)

Fig. 2. Configurations of two different attack scenarios: The zero-dynamics attack (4) requires the exact model knowledge, while the robust zero-dynamics attack (16) instead utilizes the disclosure resources (i.e., u and y). (a) Conventional zero-dynamics attack (4). (b) Robust zero-dynamics attack (16).

The first task is, using the information of the output y, to duplicate the nominal z_{n}-dynamics (18) as the form

$$\dot{z}_{\mathsf{n}}^{\mathsf{a}} = S_{\mathsf{n}} z_{\mathsf{n}}^{\mathsf{a}} + G_{\mathsf{n}} y \tag{18}$$

where $z_{\mathsf{n}}^{\mathsf{a}}(t_0)$ is chosen in $\mathcal{Z}_0$. With the auxiliary state $z_{\mathsf{n}}^{\mathsf{a}}$ and the nominal components ψ_{n}, ϕ_{n}, and g_{n}, one can rewrite the time derivative of x_ν in (1b) as

$$\begin{aligned}\dot{x}_\nu &= \psi^\top z + \phi^\top x + g(u + a) && (19a)\\ &= \psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}} + \phi_{\mathsf{n}}^\top x + g_{\mathsf{n}} u + g(a - a^\star) && (19b)\end{aligned}$$

where $a^\star \in \mathbb{R}$ is defined as

$$a^\star := \frac{1}{g}\left(-\psi^\top z + \psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}} + (\phi_{\mathsf{n}}^\top - \phi^\top)x + (g_{\mathsf{n}} - g)u \right). \tag{20}$$

Then the actual closed-loop system (5) (i.e., (1)–(3)) with the auxiliary dynamics (18) can be equivalently represented using (20) by

$$\begin{aligned}\dot{z} &= Sz + GC\chi \text{ (same as (2))}, && (21a)\\ \dot{z}_{\mathsf{n}}^{\mathsf{a}} &= S_{\mathsf{n}} z_{\mathsf{n}}^{\mathsf{a}} + G_{\mathsf{n}} C\chi, && (21b)\\ \dot{\chi} &= A_{\mathsf{n}}\chi + E_{\mathsf{n}} r + B\left(g(a - a^\star) + \psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}}\right), && (21c)\\ y &= C\chi. && (21d)\end{aligned}$$

For now, we suppose that $a^\star$ is available to the attacker, from which the attack signal a is constructed as

$$a(t) = a^\star(t), \quad \forall t \geq t_0. \tag{22}$$

It should be emphasized that under the attack $a = a^\star$, the $(z_{\mathsf{n}}^{\mathsf{a}}, \chi)$-dynamics (21b)–(21d) is exactly the same as the nominal closed-loop system (14) where the auxiliary $z_{\mathsf{n}}^{\mathsf{a}}$-dynamics disguises as the plant's internal dynamics. At the same time, the attack $a = a^\star$ in (22) leaves the real (unstable) z-dynamics (21a) decoupled from (21b)–(21d) so that the real internal state

$z(t)$ possibly diverges. The discussion so far is summarized in the following proposition.

Proposition 2: The solution $[z(t); z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t)]$ of the closed-loop system (5) under the attack (18), (20), and (22), initiated in $\mathcal{Z}_0 \times \mathcal{Z}_0 \times \mathcal{X}_0$, satisfies the following statements:

a) For almost every $[z(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)]$,[3]

$$\|z_{\mathsf{u}}(t)\| \to \infty \text{ as } t \to \infty; \tag{23}$$

b) For the solution $[z_{\mathsf{n}}; \chi_{\mathsf{n}}]$ of the nominal system (14) initiated at $[z_{\mathsf{n}}(t_0); \chi_{\mathsf{n}}(t_0)] = [z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)]$,

$$\left[z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t)\right] = \left[z_{\mathsf{n}}(t); \chi_{\mathsf{n}}(t)\right], \quad \forall t \geq t_0.$$

Moreover, $[z_{\mathsf{s}}(t); z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t)] \in \mathcal{Z}_{\mathsf{s}} \times \mathcal{Z}_{\mathsf{n}} \times \mathcal{X}$ for all $t \geq t_0$ where $\mathcal{Z}_{\mathsf{s}} \subset \mathbb{R}^{\mu_{\mathsf{s}}}$, $\mathcal{Z}_{\mathsf{n}} \subset \mathbb{R}^{\mu}$, and $\mathcal{X} \subset \mathbb{R}^{\nu+m}$ are some compact sets. ∎

Proof: The item (b) is trivial and omitted. From (b), the last statement also follows straightforwardly since $r(t)$ is uniformly bounded. Now, for (a), let us define some matrices

$$\overline{A}_{\mathsf{s}} := \begin{bmatrix} S_{\mathsf{s}} & 0 & G_{\mathsf{s}}C \\ 0 & S_{\mathsf{n}} & G_{\mathsf{n}}C \\ 0 & B\psi_{\mathsf{n}}^{\top} & A_{\mathsf{n}} \end{bmatrix}, \quad \overline{C} := \begin{bmatrix} 0_{\mu_{\mathsf{s}}}^{\top} & 0_{\mu}^{\top} & C \end{bmatrix},$$

$$\overline{B} := \begin{bmatrix} 0_{\mu_{\mathsf{s}}} \\ 0_{\mu} \\ B \end{bmatrix}, \quad \overline{E}_{\mathsf{s}} := \begin{bmatrix} 0_{\mu_{\mathsf{s}}} \\ 0_{\mu} \\ E_{\mathsf{n}} \end{bmatrix}. \tag{24}$$

Notice that $\overline{A}_{\mathsf{s}}$ is Hurwitz. By this, one obtains the *unique* solution $T \in \mathbb{R}^{\mu_{\mathsf{u}} \times (\mu_{\mathsf{s}}+\mu+\nu+m)}$ of the Sylvester equation

$$T\overline{A}_{\mathsf{s}} - S_{\mathsf{u}}T + G_{\mathsf{u}}\overline{C} = 0. \tag{25}$$

With these symbols and the coordination transformations

$$\zeta_{\mathsf{s}} := [z_{\mathsf{s}}; z_{\mathsf{n}}^{\mathsf{a}}; \chi] \quad \text{and} \quad \zeta_{\mathsf{u}} := z_{\mathsf{u}} + T\zeta_{\mathsf{s}}, \tag{26}$$

we newly represent the overall system (21) as

$$\dot{z}_{\mathsf{u}} = S_{\mathsf{u}}z_{\mathsf{u}} + G_{\mathsf{u}}\overline{C}\zeta_{\mathsf{s}} \tag{27}$$

$$\dot{\zeta}_{\mathsf{s}} = \overline{A}_{\mathsf{s}}\zeta_{\mathsf{s}} + \overline{E}_{\mathsf{s}}r + \overline{B}g(a - a^{\star}). \tag{28}$$

Differentiating ζ_{u} along with these two dynamics gives

$$\begin{aligned} \dot{\zeta}_{\mathsf{u}} &= \dot{z}_{\mathsf{u}} + T\dot{\zeta}_{\mathsf{s}} \\ &= \left(S_{\mathsf{u}}z_{\mathsf{u}} + G_{\mathsf{u}}\overline{C}\zeta_{\mathsf{s}}\right) + T\left(\overline{A}_{\mathsf{s}}\zeta_{\mathsf{s}} + \overline{E}_{\mathsf{s}}r + \overline{B}g(a - a^{\star})\right) \\ &= S_{\mathsf{u}}\zeta_{\mathsf{u}} + T\overline{E}_{\mathsf{s}}r + T\overline{B}g(a - a^{\star}). \end{aligned} \tag{29}$$

It is then clear that under $a(t) \equiv a^{\star}(t)$, the above ζ_{u}- and ζ_{s}-dynamics become

$$\dot{\zeta}_{\mathsf{s}} = \overline{A}_{\mathsf{s}}\zeta_{\mathsf{s}} + \overline{E}_{\mathsf{s}}r, \quad \text{and} \quad \dot{\zeta}_{\mathsf{u}} = S_{\mathsf{u}}\zeta_{\mathsf{u}} + T\overline{E}_{\mathsf{s}}r, \tag{30}$$

respectively, so that both are decoupled from each other. Among them, the ζ_{u}-dynamics is anti-stable (i.e., S_{u} is anti-Hurwitz) and the external signal $r(t)$ is bounded. Then, *almost all* trajectories $\zeta_{\mathsf{u}}(t)$ diverge as time goes on: more precisely, the divergence of $\zeta_{\mathsf{u}}(t)$ occurs for all admissible initial condition $\zeta_{\mathsf{u}}(t_0)$ except only one point

$$\zeta_{\mathsf{u}}(t_0) = -\int_{t_0}^{\infty} e^{-S_{\mathsf{u}}(v-t_0)}T\overline{E}_{\mathsf{s}}r(v)dv =: \zeta_{\mathsf{u},0}^{\star} \tag{31}$$

(which is well-defined because $-S_{\mathsf{u}}$ is Hurwitz). This exceptional initial condition generates the *bounded* solution of (29)

$$\zeta_{\mathsf{u}}^{\star}(t) = -\int_{t}^{\infty} e^{-S_{\mathsf{u}}(v-t)}T\overline{E}_{\mathsf{s}}r(v)dv, \quad \forall t \geq t_0. \tag{32}$$

(The readers are referred to [29], [30] for more details on the bounded solution $\zeta_{\mathsf{u}}^{\star}(t)$ for anti-stable system.)

Once $\zeta_{\mathsf{u}}(t)$ diverges as time goes on, $z_{\mathsf{u}}(t) = \zeta_{\mathsf{u}}(t) - T\zeta_{\mathsf{s}}(t)$ also does because $\zeta_{\mathsf{s}}(t)$ is bounded. Finally, one can summarize the above arguments that (23) holds if $[z(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)] \in \left(\mathcal{Z}_0 \times \mathcal{Z}_0 \times \mathcal{X}_0\right) \setminus \mathcal{L}_{\mathsf{za},0}^{\star}$ with the set

$$\begin{aligned} \mathcal{L}_{\mathsf{za},0}^{\star} := \Big\{ &\left[z(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)\right] : \\ &z_{\mathsf{u}}(t_0) + T\left[z_{\mathsf{s}}(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)\right] = \zeta_{\mathsf{u},0}^{\star} \Big\}, \end{aligned}$$

which concludes the proof. ∎

Remark 4: Item (a) of Proposition 2 highlights that unlike the conventional attack (4), the present one (22) might fail to damage the internal state $z_{\mathsf{u}}(t)$ for the specific initial conditions $[z(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)] \in \mathcal{L}_{\mathsf{za},0}^{\star}$, defined in the proof above. Worse yet, it is rarely possible to compute $\mathcal{L}_{\mathsf{za},0}^{\star}$ *a priori*, since $\zeta_{\mathsf{u},0}^{\star}$ is determined by the *future* information of the external input $r(t)$ (so $\zeta_{\mathsf{u}}^{\star}(t)$ is *non-causal*). Nonetheless, this may not be a big problem to the adversary. Indeed, the Lebesgue measure of $\mathcal{L}_{\mathsf{za},0}^{\star}$ is zero, which means that this unwanted scenario hardly occurs. ∎

Remark 5: The effect of the attack (20) at time t_0 is to replace the uncertain parameters and the state z in (19a) with the nominal ones and the state $z_{\mathsf{n}}^{\mathsf{a}}$ as in (19b). As a result, it is like replacing the zero-dynamics (21a) with (21b) at time t_0. In order to make this abrupt change as invisible as possible from the output response, it would be better to have $z_{\mathsf{n}}^{\mathsf{a}}(t_0) \approx z(t_0)$. This is possible in some situations: 1) the overall system is already in the steady state (i.e., $y(t) \approx r(t)$) before the attack is injected so that the value of z is easily guessed (at least approximately); or 2) the model uncertainty is not too large to run a state observer before the initiation of the attack using the information of y and u. If the attacker is not able to set $z_{\mathsf{n}}^{\mathsf{a}}(t_0)$ close to $z(t_0)$, then a control action of (3) may cause a transient from $t = t_0$. More discussion can be found in Section IV with some simulations. ∎

Even though the new attack policy (22) seems to resolve the considered problem directly (as in Proposition 2), there is still a huge gap between (22) and the desired attack generator (16). This is because $a^{\star}$ used in (22) is composed of uncertain parameters and unmeasured states and thus it is not obtainable in general. Yet, surprisingly, we observe that the considered situation is analogous to one in robust control theory. Indeed, $a^{\star}(t)$ represents the discrepancy between the actual and nominal

[3] Throughout this paper, the statement "a property is satisfied for almost every v (in $\mathcal{U} \subset \mathbb{R}^n$)" should be interpreted as "a property is satisfied for every $v \in \mathcal{U} \subset \mathbb{R}^n$ except those in a subset $\mathcal{U}^{\star}$ of $\mathbb{R}^n$ whose (Lebesgue) measure is zero." Note that any set $\mathcal{U}^{\star} \subset \mathbb{R}^n$ whose dimension is smaller than n has the measure zero.

plants, which has been known in the literature under the name of *lumped disturbance* (or *total disturbance*) [19], [27]. From this viewpoint, the problem of our interest can be converted into how to design a robust controller that estimates and compensates the lumped disturbance $a^\star$. Motivated by this, in the next section we will construct a *disturbance observer* [19], [20] as the robust attack generator (16). We note in advance that the disturbance observer to be presented will estimate and compensate the lumped disturbance $a^\star(t)$ (approximately but) quickly and accurately enough, from which the ideal attack policy (22) will be recovered in a practical sense.

C. Design of Robust Zero-Dynamics Attack Based on Disturbance Observer Technique

For the design of the attack generator (16), we first compute some bounds for the state variables of the overall system. Take compact sets $\hat{\mathcal{Z}}_{\mathsf{n}} \subset \mathbb{R}^\mu$ and $\hat{\mathcal{X}} \subset \mathbb{R}^\nu$ that are strictly larger than the bounded sets $\mathcal{Z}_{\mathsf{n}}$ and $\mathcal{X}$ in Proposition 2. Without loss of generality, these sets are selected large enough to satisfy $\operatorname{dist}(\partial\mathcal{Z}_{\mathsf{n}}, \partial\hat{\mathcal{Z}}_{\mathsf{n}}) > \epsilon$ and $\operatorname{dist}(\partial\mathcal{X}, \partial\hat{\mathcal{X}}) > \epsilon$ (where ϵ is given in Problem Statement). We also choose a compact set $\hat{\mathcal{Z}}_{\mathsf{s}} \supset \mathcal{Z}_{\mathsf{s}}$ such that the state trajectory $z_{\mathsf{s}}(t)$ of (1a) belongs to $\hat{\mathcal{Z}}_{\mathsf{s}}$ for all initial condition $z(t_0) \in \mathcal{Z}_0$ and for all $\chi(t) \in \hat{\mathcal{X}}$. In addition, select a positive constant $\overline{z}_{\mathsf{u}}$ to be larger than the attack specification $\underline{z}_{\mathsf{u}}$ in Problem Statement. It will be shown shortly that the state variable remains bounded as

$$\|z_{\mathsf{u}}(t)\| \le \overline{z}_{\mathsf{u}} \ \text{ and } \ \left[z_{\mathsf{s}}(t); z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t)\right] \in \hat{\mathcal{Z}}_{\mathsf{s}} \times \hat{\mathcal{Z}}_{\mathsf{n}} \times \hat{\mathcal{X}} \quad (33)$$

until the attack (16) succeeds (in the sense of item (a) in Problem Statement).

With these bounds, we consider the set

$$\mathcal{A} := \left\{a^\star \text{ in } (20) : \ \left[z_{\mathsf{u}}; z_{\mathsf{s}}; z_{\mathsf{n}}^{\mathsf{a}}; \chi\right] \text{ is bounded as } (33)\right\}$$

which contains all possible values of $a^\star$ of (20) with respect to the model uncertainty, the (bounded) state variables, and the values of $r(t)$. The set $\mathcal{A}$ is clearly bounded under the assumptions. Since computing the exact $\mathcal{A}$ can be a difficult task, we instead choose any compact set $\hat{\mathcal{A}}$ strictly larger than $\mathcal{A}$, which is enough for the design of the attack generator.

On top of that, some components to be used in the disturbance observer design are introduced below. First, let us choose a saturation function $\bar{s} : \mathbb{R} \to \mathbb{R}$ that is of $\mathcal{C}^1$ and bounded, and satisfies[4]

$$\bar{s}(\hat{a}) = \hat{a}, \ \forall \hat{a} \in \hat{\mathcal{A}} \ \text{ and } \ 0 \le \frac{\partial \bar{s}}{\partial \hat{a}}(\hat{a}) \le 1, \ \forall \hat{a} \in \mathbb{R}. \qquad (34)$$

Also, define a diagonal matrix $\Gamma(\tau) := \operatorname{diag}(\tau, \tau^2, \ldots, \tau^\nu) \in \mathbb{R}^{\nu\times\nu}$ which is invertible for any positive constant τ. (Here, τ is a design parameter to be determined in Theorem 1.) Next, choose b_i, $i = 0, \ldots, \nu-1$, such that the transfer function

$$W(s) := \frac{s^\nu + b_{\nu-1}s^{\nu-1} + \cdots + b_1 s + (\overline{g}/g_{\mathsf{n}})b_0}{s^\nu + b_{\nu-1}s^{\nu-1} + \cdots + b_1 s + (\underline{g}/g_{\mathsf{n}})b_0} \qquad (35)$$

is strictly positive real where $\underline{g}$ and $\overline{g}$ are the bounds of g in Assumption 1.

Remark 6: Such coefficients b_i can always be obtained in the following two steps. First, simply select $b_1, \ldots, b_{\nu-1}$ such that $s^{\nu-1} + b_{\nu-1}s^{\nu-2} + \cdots + b_1$ is Hurwitz. After that, choose sufficiently small $b_0 > 0$ such that the Nyquist plot of

$$G_1(s) = \frac{b_0}{s^\nu + b_{\nu-1}s^{\nu-1} + \cdots + b_1 s}$$

does not encircle the disk in the complex plane whose diameter is the real line segment between $-g_{\mathsf{n}}/\underline{g}$ and $-g_{\mathsf{n}}/\overline{g}$. For more details, see [20]. ■

Finally, following the design methodology of [28], we propose the robust zero-dynamics attack (16) as the $z_{\mathsf{n}}^{\mathsf{a}}$-dynamics (18) and

$$\dot{p}^{\mathsf{a}} = \left(\mathsf{A}_\nu - \Gamma^{-1}\beta\mathsf{C}_\nu\right)p^{\mathsf{a}} + \frac{b_0}{\tau^\nu}\mathsf{B}_\nu\left(u + a_{\mathsf{rza}} + \frac{1}{g_{\mathsf{n}}}\psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}}\right)$$

$$+ \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\left(\overline{\phi}_{\mathsf{n}} + \Gamma^{-1}\beta\right)y, \qquad (36a)$$

$$a_{\mathsf{rza}} = \bar{s}\left(\mathsf{C}_\nu p^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y\right) \qquad (36b)$$

where $p^{\mathsf{a}} \in \mathbb{R}^\nu$ and $z_{\mathsf{n}}^{\mathsf{a}} \in \mathbb{R}^\mu$ are the states of the attack generator, $\overline{\phi}_{\mathsf{n}} := [\phi_{\mathsf{n},\nu}; \cdots; \phi_{\mathsf{n},1}] \in \mathbb{R}^\nu$ and $\beta := [b_{\nu-1}; \cdots; b_0] \in \mathbb{R}^\nu$. It is noted that the output $a_{\mathsf{rza}} \in \mathbb{R}$ will serve as an estimate of $a^\star$. We take $p^{\mathsf{a}}(t_0)$ in a compact set $\mathcal{P}_0 \subset \mathbb{R}^\nu$ while $z_{\mathsf{n}}^{\mathsf{a}}(t_0)$ belongs to $\mathcal{Z}_0$. (While $\mathcal{P}_0$ can be any compact set, for example, $\mathcal{P}_0 = \{0\}$ would work, it is preferred to have $z_{\mathsf{n}}^{\mathsf{a}}(t_0) \approx z(t_0)$ as discussed in Remark 5.)

The following theorem describes our main result that the proposed attack (18) and (36) recovers the attack performance of the ideal attack policy (22) in a practical sense, while being robustly stealthy against model uncertainty.

Theorem 1: Suppose that Assumptions 1–3 hold. Then for given $\underline{z}_{\mathsf{u}} > 0$ and $\epsilon > 0$, there exists $\overline{\tau} > 0$ such that the solution $[z(t); z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t); p^{\mathsf{a}}(t)]$ of the closed-loop system (5) under the robust zero-dynamics attack $a = a_{\mathsf{rza}}$ in (18) and (36) with $\tau \in (0, \overline{\tau})$, initiated in $\mathcal{Z}_0 \times \mathcal{Z}_0 \times \mathcal{X}_0 \times \mathcal{P}_0$, satisfies the following statements:

a) For almost every $[z(t_0); z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0); p^{\mathsf{a}}(t_0)]$, there exists $t_{\mathsf{fin}} \ge t_0$ such that

$$\|z_{\mathsf{u}}(t_{\mathsf{fin}})\| > \underline{z}_{\mathsf{u}}; \qquad (37)$$

b) For the solution $[z_{\mathsf{n}}; \chi_{\mathsf{n}}]$ of the nominal system (14) initiated at $[z_{\mathsf{n}}(t_0); \chi_{\mathsf{n}}(t_0)] = [z_{\mathsf{n}}^{\mathsf{a}}(t_0); \chi(t_0)]$,

$$\left\|\left[z_{\mathsf{n}}^{\mathsf{a}}(t); \chi(t)\right] - \left[z_{\mathsf{n}}(t); \chi_{\mathsf{n}}(t)\right]\right\| < \epsilon \qquad (38)$$

for $t_0 \le t \le t_{\mathsf{fin}}$. ■

D. Proof of Theorem 1

The rest of this section is devoted to the proof of Theorem 1, which is outlined as follows. In Lemma 1, we first introduce a coordinate change for the attacker's state p^{a}, denoted by η, which serves as an error variable between a_{rza} and $a^\star$ in a sense. After that, Lemma 2 carries out the stability analysis of the η-dynamics, by which it is seen that the attack signal

[4] In other words, $\bar{s}$ is any smooth bounded function whose slope is limited by one and which is identity on $\hat{\mathcal{A}}$.

$a_{\mathsf{rza}}(t)$ quickly converges to and remains close to $a^\star(t)$ as long as $\|z_{\mathsf{u}}(t)\| \le \underline{z}_{\mathsf{u}}$ and τ is chosen sufficiently small. Then the remainder of the proof is motivated by Proposition 2 in which $a(t)$ is set as the ideal attack $a^\star(t)$ (i.e., $a(t) \equiv a^\star(t)$). Based on the Lyapunov argument, it is shown that after the convergence of $a_{\mathsf{rza}}(t)$, the actual state $[z_{\mathsf{n}}^{\mathsf{a}}(t);\chi(t)]$ behaves similar to the nominal one $[z_{\mathsf{n}}(t);\chi_{\mathsf{n}}(t)]$ until the attacker succeeds (and thus the attack remains stealthy for a while). Finally, we observe that the remaining state $z_{\mathsf{u}}(t)$ of the plant eventually encounters the set $\{z_{\mathsf{u}} : \|z_{\mathsf{u}}\| = \overline{z}_{\mathsf{u}}\}$ with a larger threshold $\overline{z}_{\mathsf{u}} > \underline{z}_{\mathsf{u}}$, for almost every initial condition of the overall system.

To this end, we represent the attacked closed-loop system into the singular perturbation form [21].

Lemma 1: With the coordinate changes (26) and

$$\eta_1 = p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y - a^\star, \tag{39a}$$

$$\eta_i = \tau^{i-1}\left(p_1^{\mathsf{a}(i-1)} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y^{(i-1)}\right), \ i = 2,\ldots,\nu, \tag{39b}$$

the overall system (5), (18), (36), and $a = a_{\mathsf{rza}}$ is transformed into the standard singular perturbation form:

(28), (29), and

$$\tau\dot\eta = \big(\mathsf{A}_\nu - \mathsf{B}_\nu\underline{\beta}^\top\big)\eta + b_0\mathsf{B}_\nu\left(\frac{g-g_{\mathsf{n}}}{g_{\mathsf{n}}}\right)\tilde a - \tau\begin{bmatrix}\dot a^\star\\ 0_{\nu-1}\end{bmatrix} \tag{40}$$

where $\underline{\beta} := [b_0;\cdots;b_{\nu-1}] \in \mathbb{R}^\nu$ and

$$\tilde a := -a_{\mathsf{rza}} + a^\star = -\overline{s}\big(C_\nu\eta + a^\star\big) + a^\star. \tag{41}$$

■

The proof of Lemma 1 is found in the Appendix.

With τ regarded as a perturbation parameter, from now on we call η the *fast* variable, while the other states the *slow* variables. The following lemma indicates that as long as the slow variables remain in the region of interest, the fast variable η approaches the *boundary layer* $\eta = 0$ during transient period.

Lemma 2: Suppose that $[z(t);z_{\mathsf{n}}^{\mathsf{a}}(t);\chi(t)]$ is bounded as in (33) for $t \ge t_0$. Then $\eta(t)$ satisfies

$$\|\eta(t)\| \le k_2 e^{-\lambda_2((t-t_0)/\tau)}\|\eta(t_0)\| + \kappa(\tau), \forall t \ge t_0, \tag{42}$$

for some positive constants k_2 and λ_2, and a class-$\mathcal{K}$ function $\kappa : \mathbb{R} \to \mathbb{R}$. ■

Proof: For convenience, define a (time-varying) nonlinear function

$$\overline{s}_{\mathsf{g}}(v,t) := \frac{g-g_{\mathsf{n}}}{g_{\mathsf{n}}}\Big[\overline{s}(v + a^\star(t)) - a^\star(t)\Big].$$

Then, under the hypothesis of the lemma, $\overline{s}_{\mathsf{g}}(0,t) \equiv 0$ and

$$\frac{\underline{g}-g_{\mathsf{n}}}{g_{\mathsf{n}}} \le \frac{\partial\overline{s}_{\mathsf{g}}}{\partial v} = \frac{g-g_{\mathsf{n}}}{g_{\mathsf{n}}}\frac{\partial\overline{s}}{\partial v} \le \frac{\overline{g}-g_{\mathsf{n}}}{g_{\mathsf{n}}}$$

by the construction of $\overline{s}$. Thus, $\overline{s}_{\mathsf{g}}(v,t)$ belongs to the *sector* $[(\underline{g}-g_{\mathsf{n}})/g_{\mathsf{n}}, (\overline{g}-g_{\mathsf{n}})/g_{\mathsf{n}}]$. Now, in the frame of scaled time $\sigma := t/\tau$, the time derivative of η with respect to σ is computed by

$$\frac{d\eta}{d\sigma} = \big(\mathsf{A}_\nu - \mathsf{B}_\nu\underline{\beta}^\top\big)\eta - b_0\mathsf{B}_\nu\overline{s}_{\mathsf{g}}(\mathsf{C}_\nu\eta,t) - \tau\begin{bmatrix}\dot a^\star\\ 0_{\nu-1}\end{bmatrix} \tag{43}$$

in which $t = \tau\sigma$.

Regarding $\tau[\dot a^\star;0_{\nu-1}]$ as a perturbation term that vanishes when $\tau = 0$, we concentrate on the stability of the unperturbed η-dynamics (i.e., (43) with $\tau = 0$), which has the form of a Lur'e-type nonlinear system

$$\frac{d\eta}{d\sigma} = \big(\mathsf{A}_\nu - \mathsf{B}_\nu\underline{\beta}^\top\big)\eta + b_0\mathsf{B}_\nu\hat u, \quad \hat y = \mathsf{C}_\nu\eta \tag{44}$$

and a nonlinearity $\hat u = -\overline{s}_{\mathsf{g}}(\hat y, t)$. It is noted that the transfer function of the linear part (44) is computed by

$$\mathsf{C}_\nu\big(sI_\nu - \mathsf{A}_\nu + \mathsf{B}_\nu\underline{\beta}^\top\big)b_0\mathsf{B}_\nu$$
$$= \frac{b_0}{s^\nu + b_{\nu-1}s^{\nu-1} + \cdots + b_0} =: G_2(s).$$

This implies that the transfer function

$$\frac{1 + ((\overline{g}-g_{\mathsf{n}})/g_{\mathsf{n}})G_2(s)}{1 + ((\underline{g}-g_{\mathsf{n}})/g_{\mathsf{n}})G_2(s)}$$

is the same as $W(s)$, and thus it is strictly positive real. The circle criterion [21, Th. 7.1] concludes that the origin of the unperturbed η-dynamics (44) is globally exponentially stable, for which a quadratic Lyapunov function $V_{\mathsf{f}}(\eta)$ exists.

Finally, one can easily obtain the lemma by differentiating the Lyapunov function V_{f} along with the perturbed system (43) and by noting that $\dot a^\star$ is a Lipschitz function of the state variables and the external inputs r and $\dot r$. ■

For further analysis, we define an error variable $[\tilde z_{\mathsf{n}}^{\mathsf{a}};\tilde\chi] := [z_{\mathsf{n}}^{\mathsf{a}} - z_{\mathsf{n}};\chi - \chi_{\mathsf{n}}]$ on the slow variables, whose time derivative is given (from (14) and (21)) by

$$\begin{bmatrix}\dot{\tilde z}_{\mathsf{n}}^{\mathsf{a}}\\ \dot{\tilde\chi}\end{bmatrix} = \begin{bmatrix}S_{\mathsf{n}} & G_{\mathsf{n}}C\\ B\psi_{\mathsf{n}}^\top & A_{\mathsf{n}}\end{bmatrix}\begin{bmatrix}\tilde z_{\mathsf{n}}^{\mathsf{a}}\\ \tilde\chi\end{bmatrix} - \begin{bmatrix}0\\ B\end{bmatrix}g\tilde a. \tag{45}$$

We remark that (45) is a stable linear system with an additional external signal $\tilde a(t)$. In particular, one has a Lyapunov function $V_{\mathsf{s}}(\tilde z_{\mathsf{n}}^{\mathsf{a}},\tilde\chi) := [\tilde z_{\mathsf{n}}^{\mathsf{a}};\tilde\chi]^\top P_{\mathsf{s}}[\tilde z_{\mathsf{n}}^{\mathsf{a}};\tilde\chi]$ where $P_{\mathsf{s}} = P_{\mathsf{s}}^\top > 0$ satisfies

$$P_{\mathsf{s}}\begin{bmatrix}S_{\mathsf{n}} & G_{\mathsf{n}}C\\ B\psi_{\mathsf{n}}^\top & A_{\mathsf{n}}\end{bmatrix} + \begin{bmatrix}S_{\mathsf{n}} & G_{\mathsf{n}}C\\ B\psi_{\mathsf{n}}^\top & A_{\mathsf{n}}\end{bmatrix}^\top P_{\mathsf{s}} = -I.$$

By differentiating V_{s} along with (45), we readily have

$$\dot V_{\mathsf{s}} < -\lambda_3 V_{\mathsf{s}} + k_3\|\tilde a\| \tag{46}$$

where λ_3 and k_3 are some positive constants. We note that the initial value of V_{s} is zero, because the nominal trajectory $[z_{\mathsf{n}}(t);\chi_{\mathsf{n}}(t)]$ of interest is initiated at the same point as the real one $[z_{\mathsf{n}}^{\mathsf{a}}(t);\chi(t)]$. In addition, due to the saturation function $\overline{s}$, $\tilde a$ has a bounded value at $t = t_0$ independent of η. From these facts, one can select $t_{\mathsf{tr}} > t_0$ sufficiently small such that

$$V_{\mathsf{s}}\big(\tilde z_{\mathsf{n}}^{\mathsf{a}}(t),\tilde\chi(t)\big) < (\epsilon^2/2)\min(\Lambda(P_{\mathsf{s}})), \quad \forall t_0 \le t \le t_{\mathsf{tr}}. \tag{47}$$

Then the inclusions in (33) are satisfied for $t_0 \le t \le t_{\mathsf{tr}}$. Notice that the inequality (38) in Theorem 1 naturally holds during the transient period $t_0 \le t \le t_{\mathsf{tr}}$. Keeping this in mind, in what follows we focus on the reduced time period $t \ge t_{\mathsf{tr}}$.

Firstly, we claim that if the slow variables are bounded as in (33) for $t \ge t_{\mathsf{tr}}$, then the fast variable $\eta(t)$ with small τ remains around the boundary layer $\eta = 0$ for that time period.

Indeed, it follows from (39) that $\eta(t_0)$ has the form of a polynomial of $1/\tau$ whose coefficients are determined by the initial conditions of the state variables. In particular, with $\tau \in (0,1)$ we have $\|\eta(t_0)\| \le \sum_{j=0}^{\nu} h_j/\tau^j$ where positive constants h_j, $j = 0, \ldots, \nu$, are independent of τ. Lemma 2 implies that

$$\|\eta(t)\| \le k_2 e^{-\lambda_2((t_{\mathsf{tr}}-t_0)/\tau)} \sum_{j=0}^{\nu} \frac{h_j}{\tau^j} + \kappa(\tau) =: \overline{\kappa}(\tau), \quad \forall t \ge t_{\mathsf{tr}} \tag{48}$$

where $\overline{\kappa} : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is continuous on τ and satisfies $\overline{\kappa}(\tau) \to 0$ as $\tau \to 0$ (because $t_{\mathsf{tr}} - t_0 > 0$). This concludes the claim. For further analysis, we particularly choose $0 < \overline{\tau}_1 < 1$ such that for all $0 < \tau < \overline{\tau}_1$, the function $\overline{\kappa}(\tau)$ satisfies

$$\overline{\kappa}(\tau) \le \min\left(\delta, (\lambda_3/k_3)\epsilon^2 \min(\Lambda(P_{\mathsf{s}}))\right) \tag{49}$$

where $\delta > 0$ is a small constant such that

$$a^\star \in \mathcal{A} \text{ and } \|\eta\| < \delta \;\Rightarrow\; \bar{s}\big(a^\star + \mathsf{C}_\nu \eta\big) = a^\star + \mathsf{C}_\nu \eta.$$

Note that such δ always exists, since the saturation level set $\hat{\mathcal{A}}$ of $\bar{s}$ is selected strictly larger than $\mathcal{A}$.

Next, we argue that for each $0 < \tau < \overline{\tau}_1$, the inequality (38) holds (and thus $a^\star(t)$ belongs to $\mathcal{A}$) until $\|z_{\mathsf{u}}(t)\| \le \bar{z}_{\mathsf{u}}$ is violated. To see this, it should be noted that if (33) holds for $t \ge t_{\mathsf{tr}}$, then it follows from (46), (48), and (49) that

$$\dot{V}_{\mathsf{s}} < -\lambda_3 \left(V_{\mathsf{s}} - \epsilon^2 \min(\Lambda(P_{\mathsf{s}}))\right).$$

This implies that the set

$$\mathcal{V} := \left\{[\tilde{z}^{\mathsf{a}}_{\mathsf{n}}; \tilde{\chi}] : V_{\mathsf{s}}(\tilde{z}^{\mathsf{a}}_{\mathsf{n}}, \tilde{\chi}) \le \epsilon^2 \min(\Lambda(P_{\mathsf{s}}))\right\}$$

is positively invariant. The proof of the argument is complete by noting that the error variable $[\tilde{z}^{\mathsf{a}}_{\mathsf{n}}(t); \tilde{\chi}(t)]$ is located inside $\mathcal{V}$ at $t = t_{\mathsf{tr}}$, and that

$$\|[\tilde{z}^{\mathsf{a}}_{\mathsf{n}}; \tilde{\chi}]\| < \epsilon \Rightarrow [z^{\mathsf{a}}_{\mathsf{n}}; \chi] = [z_{\mathsf{n}} + \tilde{z}^{\mathsf{a}}_{\mathsf{n}}; \chi_{\mathsf{n}} + \tilde{\chi}] \in \hat{\mathcal{Z}}_{\mathsf{n}} \times \hat{\mathcal{X}}.$$

At last, we complete the proof of the theorem by showing that with sufficiently small τ, there exists a finite time $t = t_{\mathrm{fin}}$ such that the partial state $z_{\mathsf{u}}(t)$ satisfies (37) for almost every $[z(t_0); z^{\mathsf{a}}_{\mathsf{n}}(t_0); \chi(t_0); p^{\mathsf{a}}(t_0)]$ in $\mathcal{Z}_0 \times \mathcal{Z}_0 \times \mathcal{X}_0 \times \mathcal{P}_0$. It is obvious from the arguments so far that as long as $\|z_{\mathsf{u}}(t)\| \le \bar{z}_{\mathsf{u}}$ and $0 < \tau < \overline{\tau}_1$, the saturation function $\bar{s}$ is inactive for $t \ge t_{\mathsf{tr}}$. Then one has $\tilde{a} = -\mathsf{C}_\nu \eta$ and $\dot{a}^\star = H_1 \eta + H_2[\zeta_{\mathsf{u}}; \zeta_{\mathsf{s}}] + H_3[r; \dot{r}]$ for some constant matrices H_i, $i = 1, 2, 3$. It follows that the overall (transformed) system (28), (29), and (40) turns out to be *linear*; in particular,

$$\tau\dot{\eta} = \big(\mathsf{A}_\nu - \mathsf{B}_\nu \underline{\beta}_{\mathsf{g}}^\top + \tau H_1\big)\eta + \tau H_2 \begin{bmatrix} \zeta_{\mathsf{u}} \\ \zeta_{\mathsf{s}} \end{bmatrix} + \tau H_3 \begin{bmatrix} r \\ \dot{r} \end{bmatrix} \tag{50}$$

where

$$\underline{\beta}_{\mathsf{g}} := \underline{\beta} + \frac{g - g_{\mathsf{n}}}{g_{\mathsf{n}}} b_0 \mathsf{C}_\nu = [(g/g_{\mathsf{n}})b_0; b_1; \cdots; b_{\nu-1}] \in \mathbb{R}^\nu.$$

For further analysis, let us consider the non-symmetric algebraic Riccati equation

$$\begin{bmatrix} \overline{E}_{\mathsf{s}} \\ 0 \end{bmatrix} + X\Big(\mathsf{A}_\nu - \mathsf{B}_\nu \underline{\beta}_{\mathsf{g}}^\top + \tau H_1\Big) - \tau \begin{bmatrix} S_{\mathsf{u}} & 0 \\ 0 & \overline{A}_{\mathsf{s}} \end{bmatrix} X - \tau R H_2 X = 0. \tag{51}$$

Here, since the matrix $\mathsf{A}_\nu - \mathsf{B}_\nu \underline{\beta}_{\mathsf{g}}^\top$ is Hurwitz, it follows from [31, Sec. 2.2] that there exists $0 < \overline{\overline{\tau}} \le \overline{\tau}_1$ such that for fixed $\tau \in (0, \overline{\overline{\tau}})$, the solution $X = X(\tau)$ of (51) is uniquely determined and its norm is bounded. Using this, we now take $\tau \in (0, \overline{\overline{\tau}})$ and define a coordinate change

$$[\hat{\zeta}_{\mathsf{u}}; \hat{\zeta}_{\mathsf{s}}] := [\zeta_{\mathsf{u}}; \zeta_{\mathsf{s}}] + \tau X \eta$$

Then with $X =: [X_{\mathsf{u}}; X_{\mathsf{s}}] \in \mathbb{R}^{\mu_{\mathsf{u}} \times \nu} \times \mathbb{R}^{(\mu_{\mathsf{s}}+n+m)\times\nu}$, it is easy to see that for $t \ge t_{\mathsf{tr}}$,

$$\begin{bmatrix} \dot{\hat{\zeta}}_{\mathsf{u}} \\ \dot{\hat{\zeta}}_{\mathsf{s}} \end{bmatrix} = \begin{bmatrix} S_{\mathsf{u}} & 0 \\ 0 & \overline{A}_{\mathsf{s}} \end{bmatrix} \begin{bmatrix} \hat{\zeta}_{\mathsf{u}} \\ \hat{\zeta}_{\mathsf{s}} \end{bmatrix} + \begin{bmatrix} T\overline{E}_{\mathsf{s}} \\ \overline{E}_{\mathsf{s}} \end{bmatrix} r + \tau \begin{bmatrix} X_{\mathsf{u}} H_3 \\ X_{\mathsf{s}} H_3 \end{bmatrix} \begin{bmatrix} r \\ \dot{r} \end{bmatrix}. \tag{52}$$

Observe that the above $\hat{\zeta}_{\mathsf{u}}$-dynamics is an anti-stable linear system with the bounded external signal r. Thus similar to the case of Proposition 2, one has

$$z_{\mathsf{u}}(t) = \hat{\zeta}_{\mathsf{u}}(t) - T\zeta_{\mathsf{s}}(t) - \tau X_{\mathsf{u}} \eta(t)$$

diverges as time goes on, as long as

$$\hat{\zeta}_{\mathsf{u}}(t_{\mathsf{tr}}) \ne -\int_{t_{\mathsf{tr}}}^{\infty} e^{-S_{\mathsf{u}}(v-t)} \left(T\overline{E}_{\mathsf{s}} r(v) + \tau X_{\mathsf{u}} H_3 \begin{bmatrix} r(v) \\ \dot{r}(v) \end{bmatrix}\right) dv =: \hat{\zeta}^\star_{\mathsf{u,tr}}.$$

To find out the exceptional case at the initial time $t = t_0$ (rather than at $t = t_{\mathsf{tr}}$), let us consider the *backward* solution of the $\hat{\zeta}_{\mathsf{u}}$-dynamics in (52). If the solution is initiated at $\hat{\zeta}_{\mathsf{u}}(t_{\mathsf{tr}}) = \hat{\zeta}^\star_{\mathsf{u,tr}}$, the corresponding value $\hat{\zeta}_{\mathsf{u}}(t_0)$, denoted by $\hat{\zeta}^\star_{\mathsf{u},0}$, is uniquely determined. Using this, we can summarize that $z_{\mathsf{u}}(t)$ must diverge if $[z(t_0); z^{\mathsf{a}}_{\mathsf{n}}(t_0); \chi(t_0); p^{\mathsf{a}}(t_0)]$ is not located in a Lebesgue measure zero set $\mathcal{L}^\star_{\mathsf{rza},0}$ on which $\hat{\zeta}_{\mathsf{u}}(t_0) = \hat{\zeta}^\star_{\mathsf{u},0}$ holds.

IV. Simulation Results: Power Generating Systems

We consider the scenario when a malicious attack enters a power generating system with a hydro turbine [32], [22], as depicted in Fig. 3. A state-space representation of the plant is given by

$$\dot{\xi}_1 = -(1/T_{\mathsf{lm}})\xi_1 + (K_{\mathsf{lm}}/T_{\mathsf{lm}})(\xi_2 - 2\xi_3), \tag{53a}$$

$$\dot{\xi}_2 = -(2/T_{\mathsf{h}})\xi_2 + (6/T_{\mathsf{h}})\xi_3, \tag{53b}$$

$$\dot{\xi}_3 = -(1/T_{\mathsf{g}})\xi_3 + (1/T_{\mathsf{g}})\big(u + a - (1/R)\xi_1\big), \tag{53c}$$

where u is the input, $y = \xi_1$ is the output, and $\xi := [\xi_1; \xi_2; \xi_3] := [\Delta f; \Delta P + 2\Delta X; \Delta X]$ is the state consisting of the frequency deviation Δf (Hz), the change in generator output ΔP (p.u.),

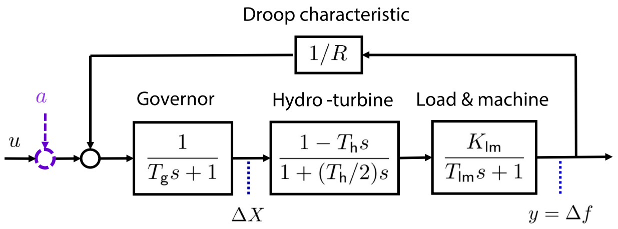

Droop characteristic
$1/R$
a
u
Governor
$\frac{1}{T_{\mathsf{g}}s+1}$
Hydro -turbine
$\frac{1-T_{\mathsf{h}}s}{1+(T_{\mathsf{h}}/2)s}$
Load & machine
$\frac{K_{\mathsf{lm}}}{T_{\mathsf{lm}}s+1}$
ΔX
$y=\Delta f$

Fig. 3. Configuration of a power generating system with a hydro turbine [22].

and the change in governor valve position ΔX (p.u.). The constants T_{lm}, T_{h}, and T_{g} indicate time constants of load and machine, hydro turbine, and governor, respectively, and R (Hz/p.u.) is the speed regulation due to the governor action. The detailed parameters of the plants are given by $K_{\mathsf{lm}}=1, T_{\mathsf{lm}}=6$, $T_{\mathsf{g}}=0.2$, and $R=0.05$, while $T_{\mathsf{h}}\in[4,6]$ is uncertain [22]. To robustly regulate the output of the uncertain plant, the control input u is generated by a (band-limited) PID-type controller $K(s)=(1.8124s^2-18.8558s+0.1523)/(0.01s^2+s)$.

For the attack design, with a suitable coordinate change

$$
\begin{aligned}
x_1 &:= \xi_1,\\
x_2 &:= -(1/T_{\mathsf{lm}})\xi_1+(K_{\mathsf{lm}}/T_{\mathsf{lm}})\xi_2-(2K_{\mathsf{lm}}/T_{\mathsf{lm}})\xi_3,\\
z &:= \xi_2+(3T_{\mathsf{lm}}/T_{\mathsf{h}})(1/K_{\mathsf{lm}})\xi_1,
\end{aligned}
$$

we transform (53) into the Byrnes–Isidori normal form (1) as follows:[5]

$$\dot z = Sz+Gy, \tag{54a}$$

$$\dot x_1 = x_2, \tag{54b}$$

$$\dot x_2 = \psi^\top z+\phi^\top x+g(u+a) \tag{54c}$$

where

$$
\begin{aligned}
\phi_1 &= -\frac{3}{T_{\mathsf{lm}}T_{\mathsf{h}}}-\frac{3}{T_{\mathsf{h}}^2}-\frac{1}{T_{\mathsf{lm}}T_{\mathsf{g}}}-\frac{3}{T_{\mathsf{lm}}T_{\mathsf{h}}}+\frac{1}{R}\frac{2K_{\mathsf{lm}}}{T_{\mathsf{lm}}}\frac{1}{T_{\mathsf{g}}},\\
\phi_2 &= -\frac{1}{T_{\mathsf{lm}}}-\frac{3}{T_{\mathsf{h}}}-\frac{1}{T_{\mathsf{g}}},\\
\psi &= \frac{K_{\mathsf{lm}}}{T_{\mathsf{lm}}}\frac{1}{T_{\mathsf{h}}}+\frac{K_{\mathsf{lm}}}{T_{\mathsf{lm}}}\frac{1}{T_{\mathsf{g}}}, g=-\frac{2K_{\mathsf{lm}}}{T_{\mathsf{lm}}}\frac{1}{T_{\mathsf{g}}},\\
S &= \frac{1}{T_{\mathsf{h}}}, \qquad G=-\frac{3}{K_{\mathsf{lm}}}\frac{1}{T_{\mathsf{h}}}-\frac{3T_{\mathsf{lm}}}{K_{\mathsf{lm}}}\frac{1}{T_{\mathsf{h}}^2}.
\end{aligned}
$$

Note that the resulting parameters ϕ, ψ, G, and S depend on T_{h}, and thus, are all uncertain, and that $S=1/T_{\mathsf{h}}>0$ so that the power generating system (53) is of non-minimum phase. It can also be seen that $\Delta f=x_1$, ΔP is a linear function of x_1 and x_2, and ΔX is a linear function of x_1, x_2, and z. Hence, with diverging $z(t)$, only $\Delta X(t)$ diverges when $x_1(t)$ and $x_2(t)$ are bounded. So, the goal of the adversary is set to enforce the valve position ΔX to become larger than 1 (p.u.) eventually, while

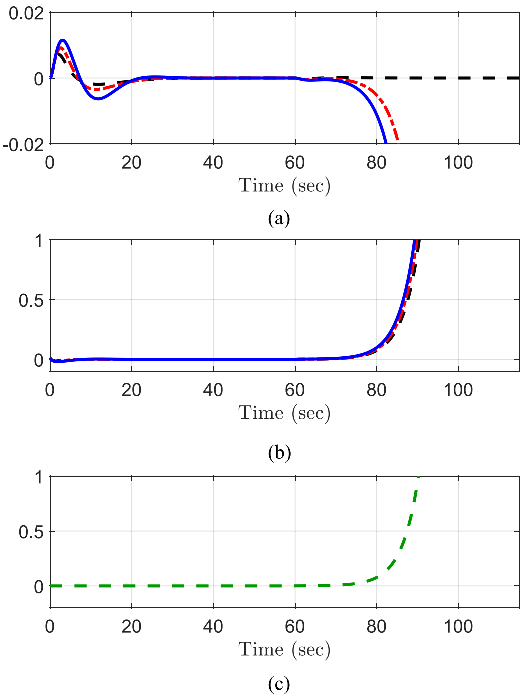


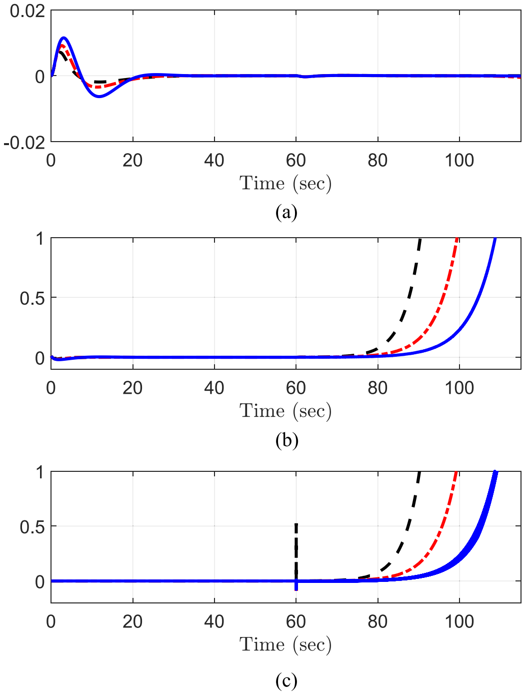
Fig. 4. Simulation results with the conventional zero-dynamics attack (11) when $T_{\mathsf{h}}=4=T_{\mathsf{h,n}}$ (black dashed), $T_{\mathsf{h}}=5$ (red dash-dotted), and $T_{\mathsf{h}}=6$ (blue solid). (a) Frequency deviation Δf (Hz). (b) Change in valve position ΔX (p.u.). (c) Conventional zero-dynamics attack a_{za} (p.u.).

the frequency deviation $\Delta f=y$ remains small as $|\Delta f|\le 0.02$ (Hz). As a result, the attack leads to overuse of water in a forebay for generating the same amount of power.

For comparison, we now construct two types of attack generator without knowledge on the value of T_{h}. One is the conventional zero-dynamics attack (11) with a nominal value $T_{\mathsf{h,n}}=4$. The other is the proposed robust zero-dynamics attack (18) and (36) designed with the same $T_{\mathsf{h,n}}$, $\overline{z}_{\mathsf{u}}=1.6$, and a saturation function $\overline{s}(\hat a)$ whose inactive region $\hat{\mathcal{A}}$ is selected as $\{\hat a : |\hat a|\le 20\,000\}$. Initial conditions are set $z^{\mathsf{a}}(t_0)=0.001$ for (11), $z_{\mathsf{n}}^{\mathsf{a}}(t_0)=-0.001$ for (18), and $p^{\mathsf{a}}(t_0)=[0;0]$ for (36). Selection for $z_{\mathsf{n}}^{\mathsf{a}}(t_0)$ is motivated by the fact that, in the steady state, the regulated output $\Delta f=y=0$ so that the steady state values for ξ and z are zero. (The effect of small mismatch between $z(t_0)$ and $z_{\mathsf{n}}^{\mathsf{a}}(t_0)$ will be observed as a small transient of $y(t)$ in the simulation result around $t=t_0$.)

Figs. 4 and 5 depict the simulation results of applying the conventional attack (11) and the proposed attacks (18) and (36) with $\tau=0.001$ to the uncertain plant (53) at the time instant $t=t_0:=60$ (sec). As shown in these figures, when there is no uncertainty, both attacks work as desired and successfully spoils the plant. However, the conventional scheme (4) immediately fails to be stealthy if it encounters the uncertain plant (Fig. 4), while the proposed attack (18) and (36) remains robust against model uncertainty (Fig. 5). It is worth noting that ΔX with the robust zero-dynamics attack diverges at different paces dependent on the value of T_{h}. Indeed, this results from the fact that the real z-dynamics under the robust zero-dynamics attack is left alone, so that the divergence of $z(t)$ depends on the unstable

[5] The detailed derivation of the normal form representation and the simulation files can be found in http://hdl.handle.net/10371/139645 and https://github.com/CDSL-GitHub/RobustZeroDynamicsAttack-Sim, respectively.

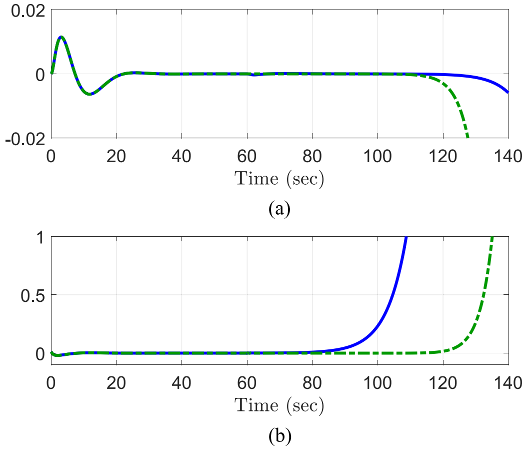


(a)

(b)

(c)

Fig. 5. Simulation results with the proposed robust zero-dynamics attack (18) and (36) when $\tau = 0.001$, $T_\mathsf{h} = 4 = T_{\mathsf{h},\mathsf{n}}$ (black dashed), $T_\mathsf{h} = 5$ (red dash-dotted), and $T_\mathsf{h} = 6$ (blue solid). (a) Frequency deviation Δf (Hz). (b) Change in valve position ΔX $(\mathrm{p.u.})$. (c) Robust zero-dynamics attack a_rza $(\mathrm{p.u.})$.



(a)

(b)

Fig. 6. Simulation results with the robust zero-dynamics attack (18) and (36) where $T_\mathsf{h} = 6$ and $\tau = 0.005$ (green dash-dotted) and $\tau = 0.001$ (blue solid). (a) Frequency deviation Δf (Hz). (b) Change in valve position ΔX_G $(\mathrm{p.u.})$.

mode of its dynamics (i.e., $S = 1/T_\mathsf{h}$). We also note that unlike the conventional zero-dynamics attack, the proposed robust zero-dynamics attack experiences a transient peak and varies by model uncertainty, in order to adjust the unknown environment by estimating and compensating $a^\star(t)$ during a short transient (Figs. 4(c) and 5(c)).

On the other hand, Fig. 6 depicts the plant's output y under the robust zero-dynamics attacks with different τ. The figure points out that for success of the attack, it is necessary for the adversary to take sufficiently small τ.

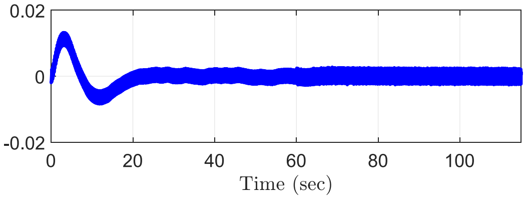


Fig. 7. Noisy output measurement $\Delta f + w$ under the robust zero-dynamics attack (18) and (36).

To investigate the presented attack further, we perform the same simulation of Fig. 5 again, with $T_\mathsf{h} = 6$ and a *noisy* measurement $y = \mathsf{C}_2 x + w$. Here, $w(t)$ is selected to have the maximum magnitude as 2×10^{-3} (Hz) and have the uniform distribution. The simulation result is depicted in Fig. 7, which indicates that the robust zero-dynamics attack still remains stealthy even in the presence of measurement noise.

V. Concluding Remarks

We have shown in this paper that fatal attacks on CPS are possible without exact system knowledge, particularly when the adversary employs robust control techniques. Specifically, we have presented a robust zero-dynamics attack that remains stealthy under model uncertainty as well as enforces the internal state of the plant to diverge.

All the results of this work indicate that more research is called for to prevent the lethal cyber attack on CPS. Since the robust zero-dynamics attack relies on the disclosure resources, a possible remedy for its prevention is to encrypt the control system [33] so that the input and output signals cannot be obtained by the adversary. One might also employ watermarking schemes [34] to detect the proposed attack.

While this paper deals with SISO systems for simplicity of explanation, extending the result of Theorem 1 to a larger class of multi-input multi-output (MIMO) plants is still possible. For instance, it is shown in [35] that the disturbance observer scheme in [20], [28] can be applied to a class of square MIMO systems with a well-defined vector relative degree. Thus by employing the modified version of the disturbance observer in [35] as an attack generator in a form similar to (18) and (36), a robust zero-dynamics attack is carried out for MIMO linear systems. A similar result also would be expected for the nonlinear systems cases, which will be addressed in future work.

Appendix

A. Proof of Lemma 1

We start by differentiating η_1 and η_i, $i = 2, \ldots, \nu - 1$, as

$$\dot\eta_1 = p_1^{\mathsf{a}(1)} - \frac{b_0}{\tau^\nu}\frac{1}{g_\mathsf{n}} y^{(1)} - \dot a^\star = \frac{1}{\tau}\eta_2 - \dot a^\star \qquad (55)$$

and

$$\dot\eta_i = \tau^{i-1}\left(p_1^{\mathsf{a}(i)} - \frac{b_0}{\tau^\nu}\frac{1}{g_\mathsf{n}} y^{(i)}\right) = \frac{1}{\tau}\eta_{i+1}, \qquad (56)$$

respectively. On the other hand, the time derivative of η_ν is given by

$$\dot{\eta}_\nu = \tau^{\nu-1}\left(p_1^{\mathsf{a}(\nu)} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y^{(\nu)}\right). \tag{57}$$

In order to compute $p_1^{\mathsf{a}(\nu)}$ in (57), from now on we show that for $i = 1, \ldots, \nu - 1$,

$$\begin{aligned} p_1^{\mathsf{a}(i)} &= p_{i+1}^{\mathsf{a}} - \sum_{k=1}^{i}\frac{b_{\nu-k}}{\tau^k}\left(p_1^{\mathsf{a}} - \frac{1}{g_{\mathsf{n}}}\frac{b_0}{\tau^\nu}y\right)^{(i-k)} \\ &\quad + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\sum_{k=1}^{i}\phi_{\mathsf{n},\nu-k+1}x_{i-k+1}. \end{aligned} \tag{58}$$

This is proved by induction. When $i = 1$, the equality (58) directly follows from (36a). Suppose that (58) is satisfied for all $i = 1, \ldots, j - 1$ and $1 \leq j \leq \nu - 1$. Then

$$\begin{aligned} p_1^{\mathsf{a}(j)} = (p_1^{\mathsf{a}(j-1)})^{(1)} &= \dot{p}_j^{\mathsf{a}} - \sum_{k=1}^{j-1}\frac{b_{\nu-k}}{\tau^k}\left(p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y\right)^{(j-k)} \\ &\quad + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\sum_{k=1}^{j-1}\phi_{\mathsf{n},\nu-k+1}\dot{x}_{(j-1)-k+1}. \end{aligned}$$

Noting that $\dot{x}_{(j-1)-k+1} = x_{j-k+1}$ for $k = 1, \ldots, j - 1$, and

$$\begin{aligned} \dot{p}_j^{\mathsf{a}} &= p_{j+1}^{\mathsf{a}} - \frac{b_{\nu-j}}{\tau^j}p_1^{\mathsf{a}} + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\left(\phi_{\mathsf{n},\nu-j+1} + \frac{b_{\nu-m}}{\tau^j}\right)y \\ &= p_{j+1}^{\mathsf{a}} - \frac{b_{\nu-j}}{\tau^j}\left(p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\right)y + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\phi_{\mathsf{n},\nu-j+1}x_1, \end{aligned}$$

one can conclude that (58) holds when $i = j$.

Now using (58) with $i = \nu - 1$ and

$$\begin{aligned} \dot{p}_\nu^{\mathsf{a}} &= -\frac{b_0}{\tau^\nu}\left(p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y\right) + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\phi_{\mathsf{n},1}x_1 \\ &\quad + \frac{b_0}{\tau^\nu}\left(u + a_{\mathsf{rza}} + \frac{1}{g_{\mathsf{n}}}\psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}}\right), \end{aligned} \tag{59}$$

one has

$$\begin{aligned} p_1^{\mathsf{a}(\nu)} &= (p_1^{\mathsf{a}(\nu-1)})^{(1)} = \dot{p}_\nu^{\mathsf{a}} - \sum_{k=1}^{\nu-1}\frac{b_{\nu-k}}{\tau^k}\left(p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y\right)^{(\nu-k)} \\ &\quad + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\sum_{k=1}^{\nu-1}\phi_{\mathsf{n},\nu-k+1}\dot{x}_{\nu-k} \\ &= -\sum_{k=1}^{\nu}\frac{b_{\nu-k}}{\tau^k}\left(p_1^{\mathsf{a}} - \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}y\right)^{(\nu-k)} \\ &\quad + \frac{b_0}{\tau^\nu}\frac{1}{g_{\mathsf{n}}}\sum_{k=1}^{\nu}\phi_{\mathsf{n},\nu-k+1}x_{\nu-k+1} + \frac{b_0}{\tau^\nu}\left(u + a_{\mathsf{rza}} + \frac{1}{g_{\mathsf{n}}}\psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}}\right). \end{aligned} \tag{60}$$

It then follows from (39), (59), and (60) that

$$\begin{aligned} \dot{\eta}_\nu &= -\frac{1}{\tau}\sum_{k=1}^{\nu}b_{\nu-k}\eta_{\nu-k+1} \\ &\quad + \frac{b_0}{\tau}\left[a_{\mathsf{rza}} - a^\star + \frac{1}{g_{\mathsf{n}}}(\phi_{\mathsf{n}}^\top x + \psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}}) + u - \frac{1}{g_{\mathsf{n}}}x_1^{(\nu)}\right] \\ &= -\frac{1}{\tau}\sum_{k=1}^{\nu}b_{\nu-k}\eta_{\nu-k+1} + \frac{b_0}{\tau}\left[a_{\mathsf{rza}} - a^\star - \frac{g}{g_{\mathsf{n}}}(a_{\mathsf{rza}} - a^\star)\right] \end{aligned} \tag{61}$$

where the latter equality comes from $x_1^{(\nu)} = \dot{x}_\nu = \psi_{\mathsf{n}}^\top z_{\mathsf{n}}^{\mathsf{a}} + \phi_{\mathsf{n}}^\top x + g_{\mathsf{n}}u_{\mathsf{c}} + g(a_{\mathsf{rza}} - a^\star)$ (since (19)) and yields the lemma.

Acknowledgment

The authors thank the anonymous reviewers and editor for their helpful suggestions on the manuscript.

References

[1] G. Park, H. Shim, C. Lee, Y. Eun, and K. H. Johansson, "When adversary encounters uncertain cyber-physical systems: Robust zero-dynamics attack with disclosure resources," in *Proc. 55th IEEE Conf. Dec. Control*, Dec. 2016, pp. 5085–5090.

[2] E. A. Lee, "Cyber physical systems: Design challenges," in *Proc. 11th IEEE Symp. Object Oriented Real-Time Distrib. Comput.*, May 2008, pp. 363–369.

[3] R. Baheti and H. Gill, "Cyber-physical systems," *The Impact of Control Technology*, pp. 161–166, 2011.

[4] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manuf. Lett.*, vol. 3, pp. 18–23, 2015.

[5] S. Gorman, "Electricity grid in U.S. penetrated by spies," *Wall Street J.*, 2009.

[6] T. Rid, "Cyber war will not take place," *J. Strategic Stud.*, vol. 35., no. 1, pp. 5–32, 2012.

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

[8] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[9] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

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

[11] C. Lee, H. Shim, and Y. Eun, "Secure and robust state estimation under sensor attacks, measurement noises, and process disturbances: Observer-based combinatorial approach," in *Proc. 2015 Eur. Control Conf.*, Jul. 2015, pp. 1872–1877.

[12] 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.

[13] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, Jul. 2014.

[14] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Proc. 50th Ann. Allerton Conf.*, Oct. 2012, pp. 1806–1813.

[15] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, Mar. 2017.

[16] C. Liu, J. Wu, C. Long, and Y. Wang, "Dynamic state recovery for cyber-physical systems under switching location attacks," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 14–22, Mar. 2017.

[17] J. Kim, G. Park, H. Shim, and Y. Eun, "Zero-stealthy attack for sampled-data control systems: The case of faster actuation than sensing," in *Proc. 55th IEEE Conf. Dec. Control*, Dec. 2016, pp. 5956–5961.

[18] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in *Proc. 2016 Amer. Control Conf.*, Jul. 2016, pp. 302–307.

[19] H. Shim, G. Park, Y. Joo, J. Back, and N. H. Jo, "Yet another tutorial of disturbance observer: Robust stabilization and recovery of nominal performance," *Control Theory Technol.*, vol. 14, no. 3, pp. 237–249, 2016, (special issue on disturbance rejection: a snapshot, a necessity, and a beginning).

[20] J. Back and H. Shim, "Adding robustness to nominal output-feedback controllers for uncertain nonlinear systems: A nonlinear version of disturbance observer," *Automatica*, vol. 44, no. 10, pp. 2528–2537, 2008.

[21] H. K. Khalil, *Nonlinear Systems*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[22] T. Wen, "Unified tuning of PID load frequency controller for power systems via IMC," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 341–350, Feb. 2010.

[23] I. Barkana, "Classical and simple adaptive control for nonminimum phase autopilot design," *J. Guidance, Control, and Dynamics*, vol. 28, no. 4, pp. 631–638, 2005.

[24] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," *IEEE Trans. Control Syst. Technol.*, vol. 8, no. 3, pp. 456–465, May 2000.

[25] G. Park, C. Lee, and H. Shim, "On stealthiness of zero-dynamics attacks against uncertain nonlinear systems: A case study with quadruple-tank process," in *Proc. Int. Symp. Math. Theory Netw. Syst.*, 2018, pp. 10–17.

[26] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 3, pp. 636–653, May 2010.

[27] J. Han, "From PID to active disturbance rejection control," *IEEE Trans. Ind. Electron.*, vol. 56, no. 3, pp. 900–906, Mar. 2009.

[28] J. Back and H. Shim, "Reduced-order implementation of disturbance observers for robust tracking of non-linear systems," *IET Control Theory Appl.*, vol. 8, no. 17, pp. 1940–1948, 2014.

[29] L. R. Hunt, G. Meyer, and R. Su, "Noncausal inverses for linear systems," *IEEE Trans. Autom. Control*, vol. 41, no. 4, pp. 608–611, Apr. 1996.

[30] Q. Jou and S. Devasia, "Preview-based stable inversion for output tracking of linear systems," *J. Dyn. Sys., Meas., Control*, vol. 121, no. 4, pp. 625–630, 1999.

[31] P. Kokotovic, H. K. Khalil, and J. O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*, SIAM, 1999.

[32] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.

[33] J. Kim *et al.*, "Encrypting controller using fully homomorphic encryption for security of cyber-physical systems," in *Proc. 6th IFAC Workshop Distrib. Estimation Control Netw. Syst.*, Sep. 2016, pp. 175–180.

[34] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst.*, vol. 35, no. 1, pp. 93–1109, Feb. 2015.

[35] J. Back and H. Shim, "An inner-loop controller guaranteeing robust transient performance for uncertain MIMO nonlinear systems," *IEEE Trans. Autom. Control*, vol. 54, no. 7, pp. 1601–1607, Jul. 2009.

Gyunghoon Park received the B.S. degree in the School of electrical and computer engineering from the Sungkyunkwan University, Seoul, South Korea, in 2011, and M.S. and Ph.D. degrees in the School of electrical engineering and computer science from the Seoul National University, Seoul, South Korea, in 2013 and in 2018, respectively.From 2018 to 2019, he was a postdoctoral researcher in Automation and System Research Institute, Seoul National University. He is currently a postdoctoral researcher in Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology. His research interests include theory and application of disturbance observer, security of cyber-physical systems, analysis of sampled-data system, and control of biped robots.

Chanhwa Lee received the B.S., M.S., and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 2008, 2010, and 2018, respectively.

From 2010 to 2012, he was an electrical engineer with Hyundai Engineering Company, Korea. Since 2018, he has been with with Hyundai Motor Company, Korea, where he is currently a senior research engineer in Research & Development Division. His research interests include security problems on cyber-physical systems, estimator design for control systems, disturbance observer technique, automotive control systems, and vehicle platooning.

Hyungbo Shim (M'93–SM'14) received the B.S., M.S., and Ph.D. degrees from Seoul National University, Seoul, Korea, and held the post-doc position at University of California, Santa Barbara, CA, USA, till 2001.

He joined Hanyang University, Seoul, Korea, in 2002. Since 2003, he has been with Seoul National University. He served as associate editor for Automatica, IEEE Trans. on Automatic Control, Int. Journal of Robust and Nonlinear Control, and European Journal of Control, and as editor for Int. Journal of Control, Automation, and Systems. He was the Program Chair of ICCAS 2014 and Vice-program Chair of IFAC World Congress 2008. His research interest includes stability analysis of nonlinear systems, observer design, disturbance observer technique, secure control systems, and synchronization.

Yongsoon Eun (M'03) received the B.A. degree in mathematics, and B.S. and M.S.E. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1992, 1994, and 1997, respectively, and Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, MI, USA, in 2003.

From 2003 to 2012, he was a Research Scientist with the Xerox Innovation Group, Webster, NY, USA, where he worked on technologies in the xerographic marking process and production inkjet printers. Since 2012, he has been with DGIST, Korea, and currently is a Professor with the Department of Information and Communication Engineering and Director of DGIST Resilient Cyber-Physical Systems Research Center. His research interests include control systems with nonlinear sensors and actuators, control of quadrotors, communication network, industry 4.0 production systems, and resilient cyber-physical systems.

Karl Henrik Johansson (F'13) received the M.Sc. and Ph.D. degrees from Lund University, Lund, Sweden.

He is currently Professor at the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and the European Control Association Council. He has received several best paper awards and other distinctions from IEEE and ACM. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar. He has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is Fellow of the Royal Swedish Academy of Engineering Sciences, and he is IEEE Distinguished Lecturer.