23rd International Symposium on Mathematical Theory of Networks and Systems
Hong Kong University of Science and Technology, Hong Kong, July 16-20, 2018

On Stealthiness of Zero-dynamics Attacks against Uncertain Nonlinear Systems: A Case Study with Quadruple-tank Process

Gyunghoon Park, Chanhwa Lee, and Hyungbo Shim*

***Abstract*— This paper studies the problem of constructing a zero-dynamics attack on "nonlinear and uncertain" cyber-physical systems being of non-minimum phase, particularly for the case of the quadruple-tank process. In most of the previous works, the zero-dynamics attack is usually designed by linearizing the nonlinear system at an operating point. As a consequence, the stealthiness of the attack may be easily violated whenever the plant has even small model uncertainty or the state trajectory under the attack moves too far from the operating point (so that the linearization is not accurate enough). Without relying on the linearization of the plant at all, in this paper we propose a nonlinear zero-dynamics attack based on the Byrnes-Isidori normal form representation. In particular, it is shown via the Lyapunov analysis that the proposed attack for the quadruple-tank process always remains stealthy until some of the tanks become empty or overflow even in the presence of small parametric uncertainty, which cannot be ensured by the existing methods. Simulation results are presented to verify the performance of the proposed attack.**

I. Introduction

A large number of control systems in these days, named *cyber-physical systems* (CPS), are often exposed to a variety of threats by malicious cyber-attacks [1]–[3]. It is in this context that the researches on possible attack scenarios to the CPS have received a lot of interests in academia. Particularly, since the physical component of the CPS can usually be modeled by differential equations, considerable model-based cyber-attack scenarios have been reported from control- and system-theoretic perspectives [4]–[10].

Among these model-based polices, zero-dynamics attack is one of the efficient and promising attack strategies to disrupt the plant with limited resources [7], [11]. Motivated by the geometric control theory, the basic idea of the attack is to inject the output-zeroing input into the actuator channel so that the attack signal can be concealed from the output measurement. By the inherent nature of the output-zeroing input, the zero-dynamics attack is effective to non-minimum phase systems. Since the pioneering work in [11], several researchers have proposed detection algorithms for the attack [11]–[14] and have sought to expand the understanding of the attack [15]. Especially, in [15], the underlying principle behind the zero-dynamics attack is reinterpreted by reformulating the attack scenario with the Byrnes-Isidori normal form. This approach in turn leads to developing another type of zero-dynamics attack that can be applied to even uncertain CPS.

It is important to note that most previous works on the zero-dynamics attack focused mainly on linear systems. This restricts applications of the zero-dynamics attack to practical problems, as cyber-physical systems often have complicated structures and are modeled by nonlinear dynamics. One simple and intuitive solution to handle the nonlinearity would be to design the zero-dynamics attack for the linearized model of the nonlinear plant around the operating point. Yet this philosophy yields the stealthiness of the attack only in a local region in most cases.

In this paper, we tackle the problem of constructing a zero-dynamics attack on "nonlinear and uncertain" cyber-physical systems being of non-minimum phase, particularly for the quadruple-tank process [16] with parametric uncertainty. By extending the discussions on the zero-dynamics attack for linear systems in [15] to the nonlinear and uncertain systems, we first transform the mathematical model of the quadruple-tank process into a Byrnes-Isidori normal form, and then employ the duplicated zero dynamics in that coordinate as an attack generator. As a result, our approach does not rely on any approximation or linearization at all. Based on the structural benefits, it is shown via the Lyapunov stability analysis that the proposed attack is stealthy in the presence of both the nonlinearity in the quadruple-tank model and small uncertainty on the plant parameter until some of the water tanks become empty or overflow.

II. System Description: Quadruple-tank Process

As a prototypical example for cyber-physical systems under cyber-attack, in this paper we consider the quadruple-tank process, introduced in [16]. Overall configuration of the plant is depicted in Fig. 1. In the figure, the control objective is to regulate the water level of the lower tanks (i.e., Tanks 1 and 2) using two pumps. It is assumed that the control input is transmitted to the pumps through data network, into which a cyber-attack is possibly injected so as to disrupt the normal operation of the system.

For a mathematical description, for $i = 1,\dots,4$ the water level of Tank i is denoted by $h_i \in \mathbb{R}_+$ [cm], whose maximum value is given by $\overline{h}_i$ [cm] due to the volume of the tank. It has been studied in the literature that in the region of interest

$$\mathscr{H} := (0,\overline{h}_1)\times(0,\overline{h}_2)\times(0,\overline{h}_3)\times(0,\overline{h}_4) \subset \mathbb{R}^4, \qquad (1)$$

This work was partly supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (2014-0-00065, Resilient Cyber-Physical Systems Research), and partially by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. NRF-2017R1E1A1A03070342).

G. Park and H. Shim are with ASRI, Department of Electrical and Computer Engineering, Seoul National University, Korea. gyunghoon.p@gmail.com, hshim@snu.ac.kr

C. Lee is with Research & Development Division, Hyundai Motor Company, Korea. chanhwa.lee@gmail.com

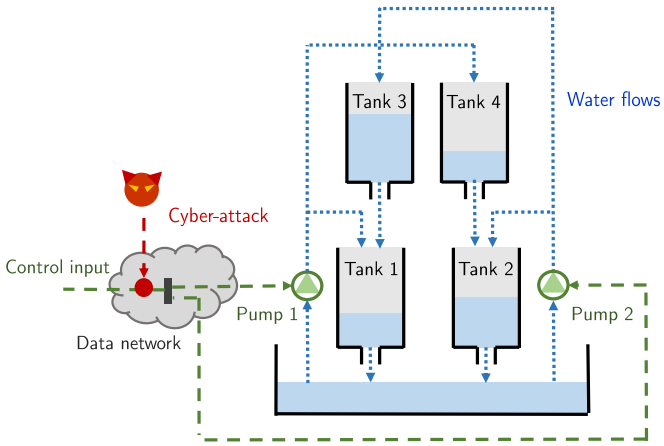


Fig. 1. Schematic diagram of quadruple-tank process under cyber-attack

with respect to h, the behavior of the quadruple-tank process in Fig. 1 can be modeled by the differential equations

$$\dot{h}_1 = -\frac{\mathsf{a}_1}{\mathsf{A}_1}\sqrt{2gh_1} + \frac{\mathsf{a}_3}{\mathsf{A}_1}\sqrt{2gh_3} + \frac{\sigma_1\mathsf{k}_1}{\mathsf{A}_1}(u_1+a_1), \tag{2a}$$

$$\dot{h}_2 = -\frac{\mathsf{a}_2}{\mathsf{A}_2}\sqrt{2gh_2} + \frac{\mathsf{a}_4}{\mathsf{A}_2}\sqrt{2gh_4} + \frac{\sigma_2\mathsf{k}_2}{\mathsf{A}_2}(u_2+a_2), \tag{2b}$$

$$\dot{h}_3 = -\frac{\mathsf{a}_3}{\mathsf{A}_3}\sqrt{2gh_3} + \frac{(1-\sigma_2)\mathsf{k}_2}{\mathsf{A}_3}(u_2+a_2), \tag{2c}$$

$$\dot{h}_4 = -\frac{\mathsf{a}_4}{\mathsf{A}_4}\sqrt{2gh_4} + \frac{(1-\sigma_1)\mathsf{k}_1}{\mathsf{A}_4}(u_1+a_1), \tag{2d}$$

and

$$y_1 = h_1, \quad y_2 = h_2. \tag{2e}$$

Here $u=(u_1,u_2)\in\mathbb{R}^2$ [V] is the control input voltage of the pumps, $y=(y_1,y_2)\in\mathbb{R}^2$ [cm] is the measurement output, and $a=(a_1,a_2)\in\mathbb{R}^2$ [V] is the attack signal. For $i=1,\dots,4$, $\mathsf{A}_i>0$ [cm^2] is the cross-section of Tank i, while $\mathsf{a}_i>0$ [cm^2] is the cross-section of the outlet hole corresponding to Tank i. In addition, k_j [cm^3/Vs], $j=1,2$, are constant gains with which $\mathsf{k}_j(u_j+a_j)$ implies the total flow passing through Pump j, and g [cm/s^2] is the acceleration of the gravity. Hence, the attack signal $a(t)$ yields unexpected variation of the water flow. We assume that the cross-sections of Tanks 1 and 3 (Tanks 2 and 4, respectively) are equal to each other, i.e., $\mathsf{a}_1=\mathsf{a}_3=:\mathsf{a}_\mathsf{L}$, $\mathsf{a}_2=\mathsf{a}_4=:\mathsf{a}_\mathsf{R}$, $\mathsf{A}_1=\mathsf{A}_3=:\mathsf{A}_\mathsf{L}$, and $\mathsf{A}_2=\mathsf{A}_4=:\mathsf{A}_\mathsf{R}$. The values of the parameters listed above are given in Table I.

On the other hand, the remaining parameters $\sigma_j\in(0,1)$, $j=1,2$, determine the proportion of the water flow that is directly injected into the lower tanks (i.e., Tanks 1 and 2). Throughout this paper, we particularly suppose that $\sigma:=(\sigma_1,\sigma_2)$ is selected by the digital controller as a fixed value, at the beginning of the system operation a priori.

In addition, it is assumed that the reference signal $r:=(r_1,r_2)$ for the output $y(t)=(y_1(t),y_2(t))$ is constant. Then one can observe that if $\sigma_1+\sigma_2\neq 1$ and $a(t)\equiv 0$ hold, then the steady-state solution $(h(t),u(t))=(h^\star_\sigma,u^\star_\sigma)$ of (2) satisfying

TABLE I
DETAILED VALUES OF PLANT PARAMETERS

Notations	Units	Values
$(\mathsf{A}_\mathsf{L},\mathsf{A}_\mathsf{R})$	[cm^2]	$(28,32)$
$(\mathsf{a}_\mathsf{L},\mathsf{a}_\mathsf{R})$	[cm^2]	$(0.071,0.057)$
$(\mathsf{k}_1,\mathsf{k}_2)$	[cm^3/Vs]	$(3.14,3.29)$
g	[cm/s^2]	981

that $(h^\star_{\sigma,1},h^\star_{\sigma,2})=(r_1,r_2)$ is *uniquely* determined. (In what follows, we often use the subscript "σ" to indicate that a constant or a variable is a function of σ.) Indeed, since $\sigma_1+\sigma_2\neq 1$,

$$u^\star_\sigma = \begin{bmatrix} u^\star_{\sigma,1}\\ u^\star_{\sigma,2}\end{bmatrix} := \begin{bmatrix} \sigma_1\mathsf{k}_1 & (1-\sigma_2)\mathsf{k}_2\\ (1-\sigma_1)\mathsf{k}_1 & \sigma_2\mathsf{k}_2\end{bmatrix}^{-1}\begin{bmatrix}\mathsf{a}_\mathsf{L}\sqrt{2gr_1}\\ \mathsf{a}_\mathsf{R}\sqrt{2gr_2}\end{bmatrix} \tag{3}$$

is well-defined. It is then easy to show that with $a(t)\equiv 0$ and $h^\star_\sigma=(h^\star_{\sigma,1},\dots,h^\star_{\sigma,4})$ where

$$h^\star_{\sigma,3} := \frac{(1-\sigma_2)^2\mathsf{k}_2^2}{2g\mathsf{a}_\mathsf{L}^2}(u^\star_{\sigma,2})^2, \quad h^\star_{\sigma,4} := \frac{(1-\sigma_1)^2\mathsf{k}_1^2}{2g\mathsf{a}_\mathsf{R}^2}(u^\star_{\sigma,1})^2, \tag{4}$$

$(h(t),u(t))=(h^\star_\sigma,u^\star_\sigma)$ satisfies the differential equation (2).

As a simple regulator for the reference command $r=(r_1,r_2)$, let us take into account a proportional integral (PI) controller

$$\dot{c} = y-r, \tag{5a}$$

$$u = u_\sigma := u^\star_\sigma + \mathsf{K}_\mathsf{p}(y-r)+\mathsf{K}_\mathsf{i}c =: u^\star_\sigma + \mathbf{K}(c,y) \tag{5b}$$

where $c\in\mathbb{R}^2$ is the state of the controller, and the gains $\mathsf{K}_\mathsf{p}\in\mathbb{R}^2$ and $\mathsf{K}_\mathsf{i}\in\mathbb{R}^2$ are selected such that the resulting controller (5) allows the output $y(t)=(y_1(t),y_2(t))$ of the closed-loop system under no attack (i.e., (2) and (5) with $a(t)\equiv 0$) to track the reference command $r=(r_1,r_2)$ exponentially. (We will come back to this point later in the next section.) For a technical reason, it is also supposed that K_i is nonsingular.

Remark 1: The main reason for specifying the controller structure as (5) is just the simplicity of the explanation, and the results to follow can be obtained similarly for generic controllers with an integral action. □

It is important to note that $\sigma_1+\sigma_2$ plays a crucial role in defining the characteristics of the plant (2). In particular, if $0<\sigma_1+\sigma_2<1$ so that most of the inlet water enters the upper tanks, then the quadruple-water tank (2) becomes of non-minimum phase. To see this, we present a coordinate change (z_σ,x) for the state h of (2) as

$$z_\sigma = \begin{bmatrix} z_{\sigma,1}\\ z_{\sigma,2}\end{bmatrix} := \begin{bmatrix} h_3-\mathsf{T}_{\sigma,2}h_2\\ h_4-\mathsf{T}_{\sigma,1}h_1\end{bmatrix}, \quad x=\begin{bmatrix}x_1\\x_2\end{bmatrix} := \begin{bmatrix}h_1\\h_2\end{bmatrix} \tag{6}$$

where $\mathsf{T}_{\sigma,1}$ and $\mathsf{T}_{\sigma,2}$ are given by

$$\mathsf{T}_{\sigma,1} := \frac{(1-\sigma_1)\mathsf{A}_\mathsf{L}}{\sigma_1\mathsf{A}_\mathsf{R}}, \quad \mathsf{T}_{\sigma,2} := \frac{(1-\sigma_2)\mathsf{A}_\mathsf{L}}{\sigma_2\mathsf{A}_\mathsf{L}}.$$

In this coordinate, the h-dynamics (2) can be represented as a *Byrnes-Isidori normal form* [17]:

$$\dot{z}_\sigma = \mathbf{H}_\sigma(z_\sigma, x), \tag{7a}$$
$$\dot{x} = \mathbf{F}_\sigma(z_\sigma, x) + \mathbf{G}_\sigma(u+a), \quad y = x, \tag{7b}$$

where the functions $\mathbf{F}_\sigma$ and $\mathbf{H}_\sigma$, and the matrix $\mathbf{G}_\sigma$ are given by

$$\mathbf{H}_\sigma(z_\sigma,x) = \begin{bmatrix}\mathbf{H}_{\sigma,1}(z_\sigma,x)\\ \mathbf{H}_{\sigma,2}(z_\sigma,x)\end{bmatrix} := \begin{bmatrix} -\frac{\sqrt{2g}\mathsf{a}_\mathsf{L}}{\mathsf{A}_\mathsf{L}}\sqrt{z_{\sigma,1}+\mathsf{T}_{\sigma,2}x_2} + \frac{\sqrt{2g}(1-\sigma_2)\mathsf{a}_\mathsf{R}}{\sigma_2\mathsf{A}_\mathsf{L}}\big(\sqrt{x_2}-\sqrt{z_{\sigma,2}+\mathsf{T}_{\sigma,1}x_1}\big)\\ -\frac{\sqrt{2g}\mathsf{a}_\mathsf{R}}{\mathsf{A}_\mathsf{R}}\sqrt{z_{\sigma,2}+\mathsf{T}_{\sigma,1}x_1} + \frac{\sqrt{2g}(1-\sigma_1)\mathsf{a}_\mathsf{L}}{\sigma_1\mathsf{A}_\mathsf{R}}\big(\sqrt{x_1}-\sqrt{z_{\sigma,1}+\mathsf{T}_{\sigma,2}x_2}\big)\end{bmatrix}, \tag{8a}$$

$$\mathbf{F}_\sigma(z_\sigma,x) = \begin{bmatrix}\mathbf{F}_{\sigma,1}(z_\sigma,x)\\ \mathbf{F}_{\sigma,2}(z_\sigma,x)\end{bmatrix} := \begin{bmatrix}\frac{\sqrt{2g}\mathsf{a}_\mathsf{L}}{\mathsf{A}_\mathsf{L}}\big(-\sqrt{x_1}+\sqrt{z_{\sigma,1}+\mathsf{T}_{\sigma,2}x_2}\big)\\ \frac{\sqrt{2g}\mathsf{a}_\mathsf{R}}{\mathsf{A}_\mathsf{R}}\big(-\sqrt{x_2}+\sqrt{z_{\sigma,2}+\mathsf{T}_{\sigma,1}x_1}\big)\end{bmatrix}, \tag{8b}$$

$$\mathbf{G}_\sigma = \mathrm{diag}\{\mathbf{g}_{\sigma,1},\mathbf{g}_{\sigma,2}\} := \mathrm{diag}\left\{\frac{\sigma_1\mathsf{k}_1}{\mathsf{A}_\mathsf{L}}, \frac{\sigma_2\mathsf{k}_2}{\mathsf{A}_\mathsf{R}}\right\}. \tag{8c}$$

With the state variable (z_σ, x), the region of interest (1) in terms of h can be rewritten by

$$\mathscr{R}_\sigma := \{(z_\sigma,x):\ (x_1,x_2,z_{\sigma,1}+\mathsf{T}_{\sigma,2}x_2, z_{\sigma,2}+\mathsf{T}_{\sigma,1}x_1)\in\mathscr{H}\}. \tag{9}$$

It is pointed out that in the new coordinate (z_σ, x), the region of interest is also dependent of σ. We also remark that by definition, the constant vectors

$$z_\sigma^\star = \begin{bmatrix}z_{\sigma,1}^\star\\ z_{\sigma,2}^\star\end{bmatrix} := \begin{bmatrix}h_{\sigma,3}^\star - \mathsf{T}_{\sigma,2}r_2\\ h_{\sigma,4}^\star - \mathsf{T}_{\sigma,1}r_1\end{bmatrix}, \quad x^\star = \begin{bmatrix}x_1^\star\\ x_2^\star\end{bmatrix} := \begin{bmatrix}r_1\\ r_2\end{bmatrix} \tag{10}$$

satisfy $\mathbf{H}_\sigma(z_\sigma^\star, x^\star) = 0$ and $\mathbf{F}_\sigma(z_\sigma^\star, x^\star) + \mathbf{G}_\sigma u_\sigma^\star = 0$; in other words, $(z_\sigma(t), x(t)) = (z_\sigma^\star, x^\star)$ is the steady-state solution of (7) without $a(t)$. Now, by applying the Lyapunov indirect method [17, Theorem 7.4] to the z_σ-dynamics (7a), one can readily obtain the following result on the non-minimum phaseness of (7).

Proposition 1: For each σ satisfying that $0<\sigma_1+\sigma_2<1$, the origin $\tilde{\delta}=0$ of the autonomous system

$$\dot{\tilde{\delta}} = \mathbf{H}_\sigma(\tilde{\delta}+z_\sigma^\star, x^\star) \tag{11}$$

is unstable. Moreover, there is a quadratic function $W_\sigma(\tilde{\delta}) = \tilde{\delta}^\top P_\sigma \tilde{\delta}$ and a constant $r_\sigma > 0$ such that $P_\sigma \in \mathbb{R}^{2\times 2}$ is a symmetric matrix and

$$\frac{\mathrm{d}W_\sigma}{\mathrm{d}\tilde{\delta}}\mathbf{H}_\sigma(\tilde{\delta}+z_\sigma^\star, x^\star) > 0, \quad \forall \tilde{\delta}\in\mathscr{D}_\sigma$$

where $\mathscr{D}_\sigma := \{\tilde{\delta}\in\mathbb{R}^2:\ \|\tilde{\delta}\|\le r_\sigma \text{ and } W_\sigma(\tilde{\delta})>0\}$. □

III. Problem Formulation

On the side of the adversary, in this paper we address the problem of constructing an attack signal $a(t)$ against the (nonlinear) quadruple-tank process (7). The particular interest here is to satisfy the following two objectives simultaneously. The primary goal of the attacker is to drive the water level $(h_3(t), h_4(t))$ of the upper tanks away from the corresponding steady-state value, and if possible, to make one or both of the upper tanks be empty or overflow eventually. At the same time, the adversary also aims to conceal such impact of the attack $a(t)$ from the measurement output $y(t)$ until the attack succeeds (so that the stealthiness of the attack is guaranteed). In other words, the water level $(h_1(t), h_2(t))$ of the lower tanks remains close to the reference command $r=(r_1,r_2)$ for a while.

Our additional concern is to deal with the problem in the presence of model uncertainty, in the sense of the following assumption.

Assumption 1: The actual value $\sigma=\sigma_\circ$ of (2) satisfies $0<\sigma_{\circ,1}+\sigma_{\circ,2}<1$ and is uncertain to the adversary. □

Remark 2: From a practical standpoint, such uncertainty on σ can take place in a variety of situations. A possibility comes from that the attacker accesses the data network in the steady-state operation of the system, whereas σ is set as $\sigma_\circ$ at the beginning of the operation a priori. On the other hand, it may also happen that the value of σ is intentionally encrypted by the digital controller before its transmission to (2), not to reveal the full model information to the adversary. □

Since the exact value $\sigma_\circ$ is not available anymore, for the model-based attack design the adversary may have to compute a set of rough estimates of $\sigma_\circ$, denoted by Γ. At this point, we suppose that such estimation is reasonable to some extent, in the sense that the predefined controller (5) can *robustly* stabilize a bundle of plants (2) with each $\sigma\in\Gamma$. To state the condition rigorously, consider the coordinate changes

$$\tilde{z}_\sigma := z_\sigma - z_\sigma^\star, \quad \tilde{x} := x - x^\star, \quad \tilde{c} := c \tag{12}$$

where $z_\sigma^\star$ and $x^\star$ are defined in (10). Then the closed-loop system (5) and (7) with $a(t)\equiv 0$ is expressed as

$$\dot{\tilde{z}}_\sigma = \mathbf{H}_\sigma(\tilde{z}_\sigma + z_\sigma^\star, \tilde{x}+x^\star), \tag{13a}$$
$$\dot{\tilde{x}} = \mathbf{F}_\sigma(\tilde{z}_\sigma + z_\sigma^\star, \tilde{x}+x^\star) + \mathbf{G}_\sigma\big(u_\sigma^\star + \mathbf{K}(\tilde{c}, \tilde{x}+x^\star)\big), \tag{13b}$$
$$\dot{\tilde{c}} = \tilde{x} + x^\star. \tag{13c}$$

Assumption 2: There exists a set

$$\Gamma \subset \{\sigma = (\sigma_1, \sigma_2) : \ 0 < \sigma_1 + \sigma_2 < 1\} \subset \mathbb{R}^2 \tag{14}$$

known to the adversary such that

(a) the actual value $\sigma_\circ$ is contained in Γ;
(b) for each $\sigma \in \Gamma$, there is a Lyapunov function $V_\sigma(\tilde{z}_\sigma, \tilde{x}, \tilde{c})$ satisfying that for all $\|(\tilde{z}_\sigma, \tilde{x}, \tilde{c})\| \le R$,

$$\mathsf{c}_1\|(\tilde{z}_\sigma, \tilde{x}, \tilde{c})\|^2 \le V_\sigma(\tilde{z}_\sigma, \tilde{x}, \tilde{c}) \le \mathsf{c}_2\|(\tilde{z}_\sigma, \tilde{x}, \tilde{c})\|^2,$$

$$\frac{\mathrm{d}V_\sigma}{\mathrm{d}(\tilde{z}_\sigma, \tilde{x}, \tilde{c})}\begin{bmatrix} \mathbf{H}_\sigma(\tilde{z}_\sigma + z_\sigma^\star, \tilde{x} + x^\star) \\ \mathbf{F}_\sigma(\tilde{z}_\sigma + z_\sigma^\star, \tilde{x} + x^\star) + \mathbf{G}_\sigma\big(u_\sigma^\star + \mathbf{K}(\tilde{c}, \tilde{x} + x^\star)\big) \\ \tilde{x} + x^\star \end{bmatrix} \le -\mathsf{c}_3\|(\tilde{z}_\sigma, \tilde{x}, \tilde{c})\|^2,$$

$$\left\| \frac{\mathrm{d}V_\sigma}{\mathrm{d}(\tilde{z}_\sigma, \tilde{x}, \tilde{c})} \right\| \le \mathsf{c}_4\|(\tilde{z}_\sigma, \tilde{x}, \tilde{c})\|$$

where $R > 0$ and $\mathsf{c}_i > 0$, $i = 1, \dots, 4$, are some constants independent of $\sigma \in \Gamma$. □

We now take a nominal value σ_n among the values in Γ. From this selection, a nominal model of (7) is carried out with the functions $\mathbf{F}_\mathsf{n} := \mathbf{F}_\sigma|_{\sigma=\sigma_\mathsf{n}}$, $\mathbf{H}_\mathsf{n} := \mathbf{H}_\sigma|_{\sigma=\sigma_\mathsf{n}}$, and the matrix $\mathbf{G}_\mathsf{n} := \mathbf{G}_\sigma|_{\sigma=\sigma_\mathsf{n}}$, which will be utilized in the attack design to follow. (For the sake of simplicity, from now one we drop the subscript "σ" of the variables and functions dependent of σ if $\sigma = \sigma_\circ$, and replace the subscript "σ" with the Sanserif font "n" if $\sigma = \sigma_\mathsf{n}$. For instance, $z := z_\sigma|_{\sigma=\sigma_\circ}$ and $z_\mathsf{n} := z_\sigma|_{\sigma=\sigma_\mathsf{n}}$.)

IV. Nonlinear Zero-dynamics Attack for Uncertain Quadruple-tank Process

As a solution to the problem presented in the previous section, we take a closer look at the *zero-dynamics attack* [11]. As studied in [11], [15], in cases when the plant to be attacked is linear and has no model uncertainty, the zero-dynamics attack can easily remain undetected from any anomaly detector, and the corresponding attack generator is the very simple form of the zero dynamics of the plant. However, when it comes to the nonlinear systems, the stealthiness is not straightforward anymore. We remind the readers that, in the previous work [11], the zero-dynamics attack has been constructed using not the exact nonlinear system directly, but its linearized model at the operating point. For instance, in our case, this indirect method results in

$$\dot{\tilde{\delta}}^\mathsf{a}_\mathrm{lza} = \mathbf{S}_\mathsf{n}\tilde{\delta}^\mathsf{a}_\mathrm{lza}, \tag{16a}$$

$$a_\mathrm{lza} = -\mathbf{G}_\mathsf{n}^{-1}\mathbf{R}_\mathsf{n}\tilde{\delta}^\mathsf{a}_\mathrm{lza} \tag{16b}$$

where

$$\mathbf{S}_\mathsf{n} := \left.\frac{\partial \mathbf{H}_\mathsf{n}(z_\mathsf{n}, x)}{\partial z_\mathsf{n}}\right|_{(z_\mathsf{n}, x) = (z_\mathsf{n}^\star, x^\star)} = \begin{bmatrix} -\frac{\sqrt{2g}\mathsf{a}_\mathsf{L}}{\mathsf{A}_\mathsf{L}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},1} + \mathsf{T}_{\mathsf{n},2}x^\star_2}} & -\frac{\sqrt{2g}(1-\sigma_{\mathsf{n},2})\mathsf{a}_\mathsf{R}}{\sigma_{\mathsf{n},2}\mathsf{A}_\mathsf{L}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},2} + \mathsf{T}_{\mathsf{n},1}x^\star_1}} \\ -\frac{\sqrt{2g}(1-\sigma_{\mathsf{n},1})\mathsf{a}_\mathsf{L}}{\sigma_{\mathsf{n},1}\mathsf{A}_\mathsf{R}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},1} + \mathsf{T}_{\mathsf{n},2}x^\star_2}} & -\frac{\sqrt{2g}\mathsf{a}_\mathsf{R}}{\mathsf{A}_\mathsf{R}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},2} + \mathsf{T}_{\mathsf{n},1}x^\star_1}} \end{bmatrix}, \tag{17a}$$

$$\mathbf{R}_\mathsf{n} := \left.\frac{\partial \mathbf{F}_\mathsf{n}(z_\mathsf{n}, x)}{\partial z_\mathsf{n}}\right|_{(z_\mathsf{n}, x) = (z_\mathsf{n}^\star, x^\star)} = \begin{bmatrix} \frac{\sqrt{2g}\mathsf{a}_\mathsf{L}}{\mathsf{A}_\mathsf{L}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},1} + \mathsf{T}_{\mathsf{n},2}x^\star_2}} & 0 \\ 0 & \frac{\sqrt{2g}\mathsf{a}_\mathsf{R}}{\mathsf{A}_\mathsf{R}}\frac{1}{2\sqrt{z^\star_{\mathsf{n},2} + \mathsf{T}^\star_{\mathsf{n},1}x^\star_1}} \end{bmatrix}. \tag{17b}$$

As a negative consequence of the linearization, the stealthiness of (16) mostly ends up being *local* near the operating point. Therefore, for the nonlinear system, how long such zero-dynamics attack can remain undetected is still questionable. It is also important to note that the (even small) model uncertainty due to σ could be another weak point of (16), as the zero-dynamics attack for the linear systems usually does.

In this section, to tackle the attack design problem in the presence of the nonlinearities in (2) and small uncertainty on σ, we propose a new type of the zero-dynamics attack. The idea is simple; we will mimic the nonlinear zero dynamics *as itself*, rather than approximate it via the linearization method, by using the Byrnes-Isidori normal form representation (7) of the quadruple-tank process. For this, we first define $\mathbf{H}^\dagger_\mathsf{n}(z_\mathsf{n}, x)$ and $\mathbf{F}^\dagger_\mathsf{n}(z_\mathsf{n}, x)$ as $\mathbf{H}_\mathsf{n}(z_\mathsf{n}, x)$ and $\mathbf{F}_\mathsf{n}(z_\mathsf{n}, x)$ with the square functions $\sqrt{\cdot}$ replaced by

$$\psi(x) := \begin{cases} \sqrt{x}, & \text{if } x \ge 0, \\ 0, & \text{otherwise.} \end{cases}$$

Unlike the original ones, the new functions $\mathbf{H}^\dagger_\mathsf{n}(z_\mathsf{n}, x)$ and $\mathbf{H}^\dagger_\mathsf{n}(z_\mathsf{n}, x)$ are globally well-defined and Lipschitz. With these functions, an extension of the conventional zero-dynamics attack for the nonlinear quadruple-tank process (2) is proposed as

$$\dot{\tilde{\delta}}^\mathsf{a}_\mathrm{nza} = \mathbf{H}^\dagger_\mathsf{n}(\tilde{\delta}^\mathsf{a}_\mathrm{nza} + z^\star_\mathsf{n}, x^\star), \tag{18a}$$

$$a_\mathrm{nza} = -\mathbf{G}_\mathsf{n}^{-1}\big(\mathbf{F}^\dagger_\mathsf{n}(\tilde{\delta}^\mathsf{a}_\mathrm{nza} + z^\star_\mathsf{n}, x^\star) - \mathbf{F}^\dagger_\mathsf{n}(z^\star_\mathsf{n}, x^\star)\big) \tag{18b}$$

where (18) is the copied version of the nominal zero dynamics. Notice that since $\sigma_\mathsf{n} \in \Gamma$, the origin of the $\tilde{\delta}^\mathsf{a}_\mathrm{nza}$-dynamics (18a) is unstable. Thus as long as the initial condition $\tilde{\delta}^\mathsf{a}_\mathrm{nza}(t^\mathsf{a}_0)$ is selected in the set $\mathscr{D}_\mathsf{n}$ (given in Proposition 1), the state trajectory $\tilde{\delta}^\mathsf{a}_\mathrm{nza}(t)$ will move away from the origin. To distinguish (18) from (16), we now call the proposed one

(18) as the *nonlinear zero-dynamics attack*, while (16) as the *linearized zero-dynamics attack*.

From now on, the stealthiness of the nonlinear zero-dynamics attack (18) is investigated. We begin by introducing a shifted steady-state value for c

$$c_{\mathsf{n}}^{\star} := \mathsf{K}_{\mathsf{i}}^{-1}(u^{\star} - u_{\mathsf{n}}^{\star}), \tag{19}$$

which will be vanished when $\sigma_{\circ}$ is exactly the same as σ_{n}, and by defining error variables

$$\tilde{z}_{\mathrm{nza}} := z - \tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} - z_{\mathsf{n}}^{\star}, \quad \tilde{c}_{\mathrm{nza}} := c - c_{\mathsf{n}}^{\star}. \tag{20}$$

Note that by definition, the controller dynamics (5) can be rewritten by

$$\begin{aligned} \dot{\tilde{c}}_{\mathrm{nza}} &= \tilde{x} + x^{\star}, && \text{(21a)} \\ u &= u^{\star} + \mathsf{K}_{\mathsf{p}}(r-y) + \mathsf{K}_{\mathsf{i}} c \\ &= u_{\mathsf{n}}^{\star} + \mathsf{K}_{\mathsf{p}}(r-y) + \mathsf{K}_{\mathsf{i}}(c - \mathsf{K}_{\mathsf{i}}^{-1}(u^{\star} - u_{\mathsf{n}}^{\star})) \\ &= u_{\mathsf{n}}^{\star} + \mathbf{K}(\tilde{c}_{\mathrm{nza}}, \tilde{x} + x^{\star}). && \text{(21b)} \end{aligned}$$

Thus one has

$$\begin{aligned} \dot{\tilde{z}}_{\mathrm{nza}} &= \mathbf{H}(z,x) - \mathbf{H}_{\mathsf{n}}^{\dagger}(\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, x^{\star}) \\ &= \mathbf{H}_{\mathsf{n}}^{\dagger}(\tilde{z}_{\mathrm{nza}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}) + \Delta_{\mathsf{z},1} + \Delta_{\mathsf{z},2} && \text{(22a)} \\ \dot{\tilde{x}} &= \mathbf{F}(z,x) + \mathbf{G}u - \mathbf{G}\mathbf{G}_{\mathsf{n}}^{-1}\big(\mathbf{F}_{\mathsf{n}}^{\dagger}(\tilde{\delta}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, x^{\star}) - \mathbf{F}_{\mathsf{n}}^{\dagger}(z_{\mathsf{n}}^{\star}, x^{\star})\big) \\ &= \mathbf{F}_{\mathsf{n}}^{\dagger}(\tilde{z}_{\mathrm{nza}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}) + \mathbf{G}_{\mathsf{n}}\big(u_{\mathsf{n}}^{\star} + \mathbf{K}(\tilde{c}_{\mathrm{nza}}, \tilde{x} + x^{\star})\big) \\ &\quad + \Delta_{\mathsf{x},1} + \Delta_{\mathsf{x},2}, && \text{(22b)} \\ \dot{\tilde{c}}_{\mathrm{nza}} &= \tilde{x} + x^{\star} && \text{(22c)} \end{aligned}$$

where the perturbation terms are given as

$$\begin{aligned} \Delta_{\mathsf{z},1} &:= \mathbf{H}(z,x) - \mathbf{H}_{\mathsf{n}}^{\dagger}(z,x), \\ \Delta_{\mathsf{z},2} &:= \mathbf{H}_{\mathsf{n}}^{\dagger}\big(\tilde{z}_{\mathrm{nza}} + \tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}\big) - \mathbf{H}_{\mathsf{n}}^{\dagger}(\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, x^{\star}) \\ &\quad - \mathbf{H}_{\mathsf{n}}^{\dagger}\big(\tilde{z}_{\mathrm{nza}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}\big) + \mathbf{H}_{\mathsf{n}}^{\dagger}(z_{\mathsf{n}}^{\star}, x^{\star}) \end{aligned}$$

and

$$\begin{aligned} \Delta_{\mathsf{x},1} &:= \mathbf{F}(z,x) - \mathbf{F}_{\mathsf{n}}^{\dagger}(z,x) + (\mathbf{G} - \mathbf{G}_{\mathsf{n}})\big(u_{\mathsf{n}}^{\star} + \mathbf{K}(\tilde{c}_{\mathrm{nza}}, \tilde{x} + x^{\star})\big), \\ \Delta_{\mathsf{x},2} &:= \mathbf{F}_{\mathsf{n}}^{\dagger}\big(\tilde{z}_{\mathrm{nza}} + \tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}\big) - \mathbf{F}_{\mathsf{n}}^{\dagger}(\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}} + z_{\mathsf{n}}^{\star}, x^{\star}) \\ &\quad - \mathbf{F}_{\mathsf{n}}^{\dagger}\big(\tilde{z}_{\mathrm{nza}} + z_{\mathsf{n}}^{\star}, \tilde{x} + x^{\star}\big) + \mathbf{F}_{\mathsf{n}}^{\dagger}(z_{\mathsf{n}}^{\star}, x^{\star}). \end{aligned}$$

It should be noted that as long as the plant's state $(z(t), x(t))$ remains in the bounded set $\mathscr{R}$ (i.e., region of interest), we have

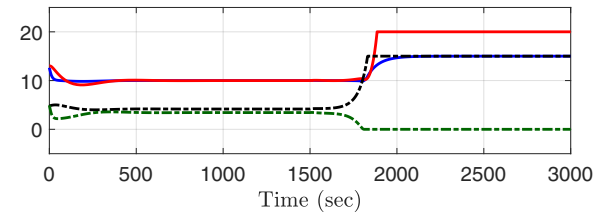
$$\begin{aligned} \|\Delta_1\| &:= \|(\Delta_{\mathsf{z},1}, \Delta_{\mathsf{x},1})\| \\ &\le M_{\Delta,1}\|\sigma_{\circ} - \sigma_{\mathsf{n}}\| + L_{\Delta,1}\|\sigma_{\circ} - \sigma_{\mathsf{n}}\|\|(\tilde{x}, \tilde{c}_{\mathrm{nza}})\|, && \text{(23a)} \\ \|\Delta_2\| &:= \|(\Delta_{\mathsf{z},2}, \Delta_{\mathsf{x},2})\| \le L_{\Delta,2}\|(\tilde{z}_{\mathrm{nza}}, \tilde{x})\| && \text{(23b)} \end{aligned}$$

for some positive constants $M_{\Delta,1}$, $L_{\Delta,1}$, and $L_{\Delta,2}$.

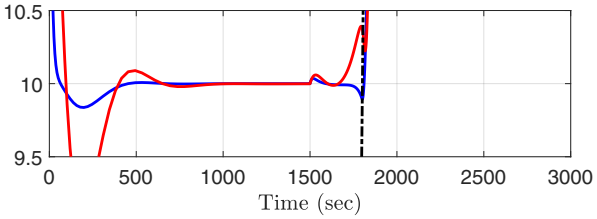
We also note that by definition, there exists $M^{\star} > 0$ such that

$$\|(z^{\star} - z_{\mathsf{n}}^{\star}, 0, -c_{\mathsf{n}}^{\star})\| \le M^{\star}\|\sigma_{\circ} - \sigma_{\mathsf{n}}\|. \tag{24}$$

The following theorem describes our main result, which indicates that if σ_{n} is selected close enough to $\sigma_{\circ}$, then the nonlinear zero-dynamics attack (18) remains stealthy until the attack disrupts the plant as much as desired, under a few more assumptions.



(a) Water level $h(t)$



(b) Enlargement

Fig. 2. Water level $h(t)$ under linearized zero-dynamics attack (16) without uncertainty on σ: $h_1(t)$ (blue solid), $h_2(t)$ (red solid), $h_3(t)$ (black dash-dotted), and $h_4(t)$ (green dash-dotted)

Theorem 1: Suppose that Assumptions 1 and 2 hold. Then for given $0 < \varepsilon < R$, the closed-loop system (5) and (7) under the nonlinear zero-dynamics attack $a = a_{\mathrm{nza}}$ with (18) satisfies

$$\|y(t) - r\| < \varepsilon, \tag{25a}$$

$$\|z(t) - \tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t) - z_{\mathsf{n}}^{\star}\| < \varepsilon \tag{25b}$$

as long as $(z(t), x(t))$ remains in the region of interest $\mathscr{R}$, if the following conditions hold:

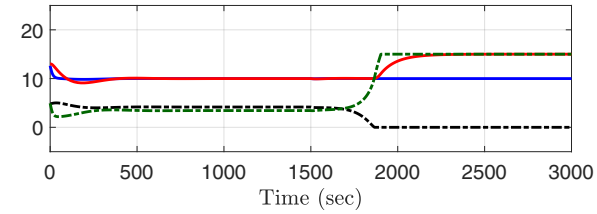
(a) $\mathsf{c}_3 - L_{\Delta,2}\mathsf{c}_4 =: \alpha_{\mathrm{nza}} > 0$;
(b) $\|(z(t_0^{\mathsf{a}}), x(t_0^{\mathsf{a}}), c(t_0^{\mathsf{a}})) - (z^{\star}, x^{\star}, 0)\| < (1/\sqrt{\mathsf{c}_2})(\varepsilon/3)$;
(c) $\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t_0^{\mathsf{a}}) \in \mathscr{D}_{\mathsf{n}}$ and $\|\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t_0^{\mathsf{a}})\| < (1/\sqrt{\mathsf{c}_2})(\varepsilon/3)$;
(d) the nominal value σ_{n} satisfies

$$\|\sigma_{\circ} - \sigma_{\mathsf{n}}\| \le \min\left\{\sqrt{\frac{\mathsf{c}_1}{\mathsf{c}_2}}\frac{\alpha_{\mathrm{nza}}}{M_{\Delta,1}\mathsf{c}_4}\varepsilon, \frac{1}{M^{\star}\sqrt{\mathsf{c}_2}}\frac{\varepsilon}{3}, \frac{\alpha_{\mathrm{nza}}}{2L_{\Delta,2}\mathsf{c}_4}\right\}.$$

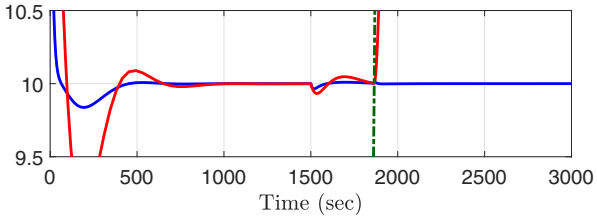
□

Proof: The theorem is proved via the Lyapunov stability analysis for the perturbed system (22). By differentiating the Lyapunov function candidate V_{n} in Assumption 2 along with (22), one has

$$\begin{aligned} \dot{V}_{\mathsf{n}} &= \frac{\mathrm{d}V_{\mathsf{n}}}{\mathrm{d}(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})}(\dot{\tilde{z}}_{\mathrm{nza}}, \dot{\tilde{x}}, \dot{\tilde{c}}_{\mathrm{nza}}) \\ &\le -\mathsf{c}_3\|(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})\|^2 + \left\|\frac{\mathrm{d}V_{\mathsf{n}}}{\mathrm{d}(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})}\right\|\big(\|\Delta_1\| + \|\Delta_2\|\big) \\ &\le -\big(\mathsf{c}_3 - L_{\Delta,2}\mathsf{c}_4 - L_{\Delta,2}\mathsf{c}_4\|\sigma_{\circ} - \sigma_{\mathsf{n}}\|\big)\|(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})\|^2 \\ &\quad + \mathsf{c}_4 M_{\Delta,1}\|\sigma_{\circ} - \sigma_{\mathsf{n}}\|\|(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})\| \\ &\le -\frac{\alpha_{\mathrm{nza}}}{2}\|(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})\| \\ &\quad \times \left(\|(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}})\| - \frac{2\mathsf{c}_4 M_{\Delta,1}}{\alpha_{\mathrm{nza}}}\|\sigma_{\circ} - \sigma_{\mathsf{n}}\|\right). \end{aligned}$$

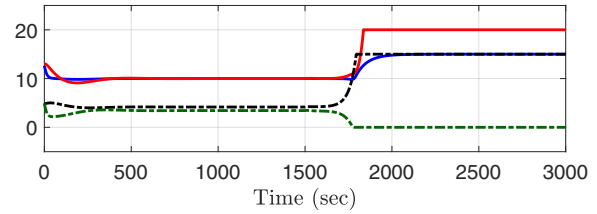


(a) Water level $h(t)$

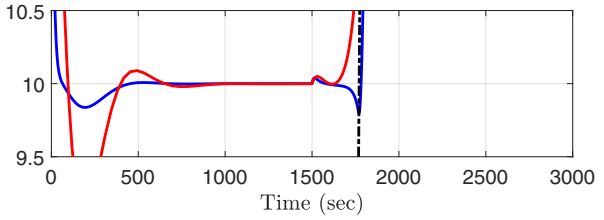


(b) Enlargement

Fig. 3. Water level $h(t)$ under nonlinear zero-dynamics attack (18) without uncertainty on σ: $h_1(t)$ (blue solid), $h_2(t)$ (red solid), $h_3(t)$ (black dash-dotted), and $h_4(t)$ (green dash-dotted)



(a) Water level $h(t)$



(b) Enlargement

Fig. 4. Water level $h(t)$ under linearized zero-dynamics attack (16) with uncertainty on σ: $h_1(t)$ (blue solid), $h_2(t)$ (red solid), $h_3(t)$ (black dash-dotted), and $h_4(t)$ (green dash-dotted)

By the assumptions, it is easy to see that the set

$$\mathscr{V} := \Bigg\{ (\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}}) : V_{\mathsf{n}}(\tilde{z}_{\mathrm{nza}}, \tilde{x}, \tilde{c}_{\mathrm{nza}}) \le \mathsf{c}_2 \frac{4\mathsf{c}_4^2 M_{\Delta,1}^2}{\alpha_{\mathrm{nza}}^2} \|\sigma_\circ - \sigma_{\mathsf{n}}\|^2 < \mathsf{c}_1 \varepsilon^2 \Bigg\}$$

is invariant. In addition, we also have

$$\begin{aligned} &\|(\tilde{z}_{\mathrm{nza}}(t_0^{\mathsf{a}}), \tilde{x}(t_0^{\mathsf{a}}), \tilde{c}_{\mathrm{nza}}(t_0^{\mathsf{a}}))\| \\ &\le \|(z(t_0^{\mathsf{a}}), x(t_0^{\mathsf{a}}), c(t_0^{\mathsf{a}}) - (z^\star, x^\star, 0)\| + \|\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t_0^{\mathsf{a}})\| \\ &\quad + \|(z^\star - z_{\mathsf{n}}^\star, 0, -c_{\mathsf{n}}^\star)\| \le \frac{1}{\sqrt{\mathsf{c}_2}} \varepsilon, \end{aligned}$$

by which the initial condition $(\tilde{z}_{\mathrm{nza}}(t_0^{\mathsf{a}}), \tilde{x}(t_0^{\mathsf{a}}), \tilde{c}_{\mathrm{nza}}(t_0^{\mathsf{a}}))$ belongs to the invariant set $\mathscr{V}$. This concludes the proof. ■

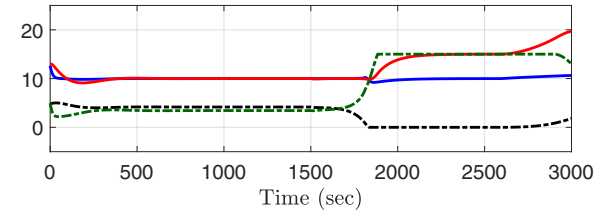
Some remarkable points in Theorem 1 are listed below.

- The inequality (25a) implies that the nonlinear zero-dynamics attack (18) remains stealthy in a practical sense. Moreover, the stealthiness of the proposed attack is *surely* guaranteed until the success of the attack, whereas the linearized zero-dynamics attack (16) cannot ensure anything about the stealthiness. From this perspective, it can be concluded that the nonlinear zero-dynamics attack (18) is more threatening than the traditional one in the cases of nonlinear systems.
- On the other hand, the second inequality (25b) illustrates how the nonlinear zero-dynamics attack (18) disrupts the internal state $z(t)$ of the plant (7). Indeed, since the initial condition $\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t_0^{\mathsf{a}})$ of the attack is set as $\mathscr{D}_{\mathsf{n}}$, the attacker's state $\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t)$ goes away from the neighborhood of the origin, at least for a while. Then for small $\varepsilon > 0$, the actual state $z(t)$ behaves as $\tilde{\delta}_{\mathrm{nza}}^{\mathsf{a}}(t) + z_{\mathsf{n}}^\star$, which must escape the ideal position $z = z_{\mathsf{n}}^\star$. It is further noted that, as the deviation of $z(t)$ gets larger, that of the water level $(h_3(t), h_4(t))$ of the upper tanks must also do. This is because the output $(h_1(t), h_2(t))$ will remain around the reference r by (25a).
- We point out that if the functions $\mathbf{H}_{\mathsf{n}}^\dagger$ and $\mathbf{F}_{\mathsf{n}}^\dagger$ are linear, then the vanishing perturbation terms $\Delta_{\mathsf{x},2}$ and $\Delta_{\mathsf{z},2}$ are naturally zero (so that $L_{\Delta,2} = 0$), and thus Item (a) in the theorem seems meaningless. In other words, Item (a) of the theorem is required in order to struggle with the nonlinearity of the plant. To achieve the constraint, the adversary may have to rely more on the inherent characteristics of the existing controller (5), which is not necessary for the linear system cases.
- On the other hand, Item (b) highlights that in order to remain stealthy, it is suggested to initiate the attack signal $a(t)$ during the steady-state operation of the system. This is a reasonable requirement to most adversaries.
- Due to the Lyapunov analysis, an endurable quantity of model uncertainty for the stealthiness of the nonlinear zero-dynamics attack (18) can be explicitly obtained, as stated in Item (d).

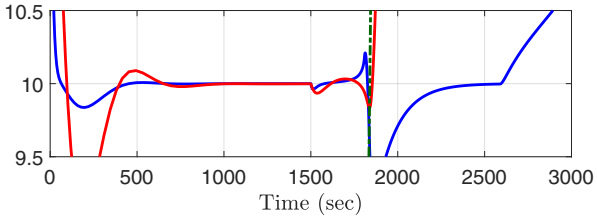
V. Simulation Results

In this section, some simulation results are presented to compare the two types of the zero-dynamics attack, (16) and (18). In the simulation, we set $h(0) = (12.6, 13, 4.8, 4.9)$, $c(0) = 0$, and $t_0^{\mathsf{a}} = 1500$ [s]. The gain matrices K_{p} and K_{i} are selected as diagonal matrices $\mathsf{K}_{\mathsf{p}} = \mathrm{diag}(0.75, -0.06)$ and $\mathsf{K}_{\mathsf{i}} = \mathrm{diag}(0.0068, -0.00027)$, by which the controller (5) is of the decentralized form (as in [16]). The actual value $\sigma_\circ$ is given by $\sigma_\circ = (0.43, 0.34)$. For a fair comparison, the initial conditions of two attack generators are set as the same value.

Figs. 2 and 3 depicts the simulation results when there is no uncertainty on σ (i.e., $\sigma_\circ = \sigma_{\mathsf{n}}$). It can be seen in these figures that as $z(t)$ diverges from the steady-state value,

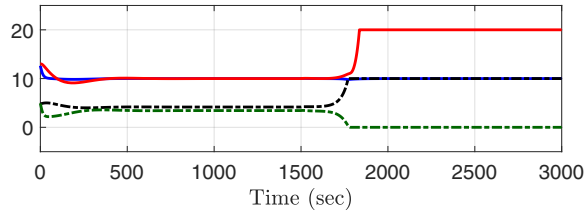


(a) Water level $h(t)$

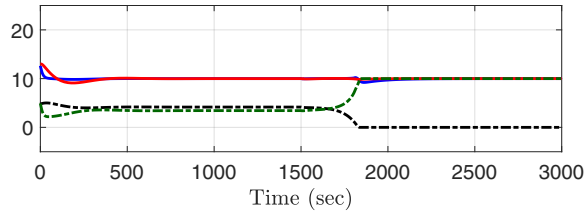


(b) Enlargement

Fig. 5. Water level $h(t)$ under nonlinear zero-dynamics attack (18) with uncertainty on σ: $h_1(t)$ (blue solid), $h_2(t)$ (red solid), $h_3(t)$ (black dash-dotted), and $h_4(t)$ (green dash-dotted)



(a) Linearized zero-dynamics attack (16)



(b) Nonlinear zero-dynamics attack (18)

Fig. 6. Water level $h(t)$ under zero-dynamics attacks with uncertainty on σ, in cases when (r_1, r_2) is equal to $(\bar{h}_3, \bar{h}_4)$: $h_1(t)$ (blue solid), $h_2(t)$ (red solid), $h_3(t)$ (black dash-dotted), and $h_4(t)$ (green dash-dotted)

the impact of the linearized zero-dynamics attack (16) gets revealed more and more in the output channel. On the other hand, the proposed zero-dynamics attack (18) seems almost stealthy until $h_3(t)$ and $h_4(t)$ touches the boundary of the region of interest. Similar conclusions can be found in Figs. 4 and 5, in which $\sigma_\circ = (0.47, 0.30) \neq \sigma_\mathrm{n}$.

We further remark that unlike the linearized zero-dynamics attack, the proposed attack may possibly be undetected even after the upper tanks become empty or overflow, under particular conditions on the system characteristics. For instance, Fig. 6 shows the scenario when $(\bar{h}_3, \bar{h}_4)$ is equal to $(r_1, r_2) = (10, 10)$, in which the impact of the nonlinear zero-dynamics attack is rarely observed.

VI. Conclusion

In this paper, we have investigated the stealthiness of the zero-dynamics attack for the quadruple-tank process, as a prototypical example for nonlinear, multi-input multi-output, and non-minimum phase systems. A typical way of implementing the zero-dynamics attack scheme for a nonlinear cyber-physical system is to linearize the dynamics of the plant and then to construct a linear attack generator; yet this is not a complete solution to the attackers because the approximation error of the linearization becomes significantly large as the internal state diverges from its initial location, by which the stealthiness of the attack is readily violated. Moving away from the conventional approach, we present a nonlinear version of the zero-dynamics attack based on the Byrnes-Isidori normal form. It has been seen from the Lyapunov analysis that the proposed zero-dynamics attack can remain stealthy in the presence of nonlinearity and even small parametric uncertainty in a practical sense, until some of the water tanks become empty or overflow. This work highlights that more threatening attack strategies would be possible as long as adversaries utilize nonlinear system theory in their attack designs.

References

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

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

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

[4] 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.

[5] 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, 2013.

[6] 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, 2014.

[7] 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, 2015.

[8] 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 European Control Conf.*, July 2015, pp. 1872–1877.

[9] 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*, on-line available.

[10] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Ann. Allerton Conf.*, Sept. 2009, pp. 911–918.

[11] 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.

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

[13] J. Back, J. Kim, C. Lee, G. Park, and H. Shim, "Enhancement of security against zero dynamics attack via generalized hold," in *Proc. 56th IEEE Conf. Dec. Control*, Dec. 2017.

[14] M. Naghnaeian, N. Hirzallah, and P. G. Voulgaris, Dual rate control for security in cyber-physical systems, in *Proc. 54th IEEE Conf. Dec. Control*, Dec. 2015, pp. 14151420.

[15] 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.
[16] K. J. Johansson, "The quadruple-tank process: A multivariable labratory process with an adjustable zero," *IEEE Trans. Control Sys. Technol.*
[17] H. K. Khalil, *Nonlinear Systems* (3rd ed.), Prentice Hall, 1996.