2025 23rd European Control Conference (ECC)
June 24-27, 2025. Thessaloniki, Greece

Zero Dynamics Attacks Subject to Actuator Saturation: A Constrained Optimization Approach

Bart Wolleswinkel, Manuel Mazo Jr., and Riccardo Ferrari

***Abstract*— Zero dynamics attacks (ZDAs) have received considerable attention in the control systems literature, as they can be disruptive while being almost virtually to detect from the measured output of the plant. However, as ZDAs require an unbounded input sequence, the effect of physical constraints on the actuators, in the form of saturation, must be taken into account. In this work, we show that conventional methods for constructing ZDAs, when subject to input saturation, can make these attacks no longer disruptive, stealthy, or both. While this might imply that some systems are safe from ZDAs, we introduced a new attack called a relaxed ZDA, which can be disruptive and practically stealthy even under input constraints. For the construction of relaxed ZDAs, we propose a method that involves solving an optimization problem offline. We demonstrate the versatility of the proposed method and show it succeeds where conventional ZDAs fall short by means of an illustrative example on a cyber-physical system (CPS).**

I. Introduction

The past decades have seen an increase in the use of digital technologies in controlling physical processes, including critical infrastructure, leading to cyber-physical system (CPS). Aside from the use of digital controllers in control automation, the use of digital (and often wireless) communication networks has also seen a strong increase. In these networked control systems (NCSs), the plant and the controller are physically non-collocated, and signals are exchanged over a network.

Whilst NCSs boast several advantages, they also pose new challenges, in particular, their vulnerability to cyberattacks. In the past decades, we have seen cyberattacks specifically targeting industrial control systems (ICSs) such as the Stuxnet computer worm, the Maroochy water breach, and ongoing attacks on the Ukrainian power grid [1]. These incidents have demonstrated the need for enhanced cybersecurity in control systems, establishing the field of secure control [2].

Examples of attacks on control systems include denial-of-service (DoS) attacks [3], which aim to disrupt the signals sent over the communications network, and *deception attacks* [4], which aim to inflict damage to the physical process whilst actively deceiving the control system. ZDAs, which we discuss here, are model-based attacks targeting non-minimum phase (NMP) plants and belong to the latter.

ZDAs have been shown, in theory, to be a threat to a variety of systems, including linear time-invariant (LTI) systems [5], [6], sampled-data systems [7], uncertain [8] and nonlinear system [9], and non-uniformly sampled systems [10]. Their interaction with noise [8] and quantization [11] has also been studied, and several countermeasures have also been proposed [5], [12], [7].

One property intrinsic to ZDAs is that the input sequence injected into the plant becomes unbounded, as ZDAs target unstable zeros. In all the above-mentioned work, and most of the work on secure control [13], the effect of input saturation is not taken into consideration. However, for these attacks to have a destructive effect on the physical process, they need to be able to be executed by the actuators, which are always subject to physical limitations [13]. This restricts the trajectories that actuators can induce, as also evident in an experimental ZDA carried out in [5].

As such, this might imply that, in practice, saturation impairs the disruptive effect of a ZDA, which begs the question of whether ZDAs should be feared at all. Here, we show that by sacrificing some stealthiness, we show that what we call *relaxed* ZDAs can still be disruptive. By remaining stealthy in a practical sense, the detection of the attack is sufficiently delayed until fatal damage is incurred by the plant. We furthermore provide a framework to construct relaxed ZDAs, as well as conventional ZDAs, as the solution to an optimization problem.

Our work is similar to that of [13], where they consider posing artificial limits on the actuators to constrain attacker capabilities; here, we do not impose further restrictions and instead concern ourselves with the stealthiness of attacks. To the best of the authors' knowledge, this is the first time the feasibility of ZDAs under actuator saturation has been explicitly taken into consideration.

The contributions of our work are threefold. First, we formally show that three methods commonly used in literature to construct ZDAs lead to identical attack sequences. Second, we demonstrate that conventional ZDAs might not be successful when input saturation is present. Thirdly, our main result is that we introduce the notion of relaxed ZDAs, and we present a new optimization-based method for constructing both conventional and relaxed ZDAs.

Notation: Let col stack its operands vertically, such that $\mathrm{col}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_m) = \begin{bmatrix} \boldsymbol{v}_1^\mathsf{T} & \cdots & \boldsymbol{v}_m^\mathsf{T} \end{bmatrix}^\mathsf{T}$. Given a vector $\boldsymbol{v} \in \mathbb{R}^n$, let $[\boldsymbol{v}]_i \in \mathbb{R}$ denote the i-th entry, and for a matrix $\boldsymbol{A}$, let $\boldsymbol{a} = [\boldsymbol{A}]_{\bullet j}$ denote the j-th column. Given a complex number $z = a + b\mathfrak{i} \in \mathbb{C}$, let $\overline{z} = a - b\mathfrak{i}$ denote its complex conjugate. For a square matrix $\boldsymbol{A}$ and subspace V, let the spectrum $\sigma(\boldsymbol{A})$ denote the multiset[1] of eigenvalues of $\boldsymbol{A}$, and let $\sigma(\boldsymbol{A} \,|\, V)$ denote the spectrum whose corresponding eigenvectors are contained in V. Given two sets $\mathbb{V}, \mathbb{W} \subset \mathbb{R}^n$,

This work has been partially supported by the EU Horizon program through the project TWAIN, grant id 101122194.

The authors are with the Delft Center for Systems and Control (DCSC), Mechanical Engineering Faculty, Delft University of Technology, Delft, The Netherlands (email:{b.wolleswinkel,r.ferrari}@tudelft.nl).

let $\mathbb{V} \oplus \mathbb{W} = \{w+v \,|\, v \in \mathbb{W}, v \in \mathbb{V}\}$ denote their Minkowski sum. We use $\boldsymbol{x}_{a:b} = (\boldsymbol{x}_a, \ldots, \boldsymbol{x}_b)$, $a, b \in \mathbb{N}_0$ to denote a sequence of vectors.

II. System description

We consider the following discrete-time LTI plants, with $k \in \mathbb{N}_0$, and individually-bounded inputs:

$$\mathcal{P}: \quad \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k, \tag{1a}$$

$$\boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{v}_k, \tag{1b}$$

with state $\boldsymbol{x}_k \in \mathbb{R}^{n_\mathrm{x}}$, physical input $\boldsymbol{u}_k \in \mathbb{R}^{n_\mathrm{u}}$, measurement $\boldsymbol{y}_k \in \mathbb{R}^{n_\mathrm{y}}$, measurement noise $\boldsymbol{v}_k \in \mathbb{R}^{n_\mathrm{y}}$, and matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ all of appropriate dimensions. The actuators are subject to saturation such that, for all $i \in \{1, \ldots, n_\mathrm{u}\}$,

$$[\boldsymbol{u}_k]_i = \mathrm{sat}([\boldsymbol{c}_k + \boldsymbol{a}_k]_i \,; \underline{u}_i, \bar{u}_i), \tag{2}$$

where $\boldsymbol{c}_k \in \mathbb{R}^{n_\mathrm{u}}$ is the control input, $\boldsymbol{a}_k \in \mathbb{R}^{n_\mathrm{u}}$ is the malicious input (see Section II-A), and sat is a the saturation function given by $\mathrm{sat}(u \,; \underline{u}, \bar{u}) = \min\{\bar{u}, \max\{\underline{u}, u\}\}$, where $\underline{u}, \bar{u}$ denote the lower and upper bound, respectively. We consider *square* multiple-input and multiple-output (MIMO) systems, and make the following standard assumption:

Assumption 1. *The pairs $(\boldsymbol{A}, \boldsymbol{B})$ and $(\boldsymbol{A}, \boldsymbol{C})$ are controllable and observable, respectively.* ◊

The plant $\mathcal{P}$ and digital controller $\mathcal{C}$ are physically non-collocated, and transmission over a communications channel is done periodically (see Fig. 1). The objective of the controller $\mathcal{C}$ is (output) regulation, whilst an adversary $\mathcal{A}$ designs a malicious input $\boldsymbol{a}_k$, which we discuss next.

A. Adversary model

We assume an adversary $\mathcal{A}$ is present within the network layer between the controller and the plant [2], trying to inflict physical damage to the control system. Following the framework of [6], the adversary has *disruption capabilities* over the control inputs are sent, but not the sensor measurements. Furthermore, the adversary does not necessarily have *disclosure resources* of either $\boldsymbol{c}_k$ or $\boldsymbol{y}_k$.

The control system operates within a (polytopic) *safe set* $\mathbb{X}_\mathrm{safe} \subset \mathbb{R}^{n_\mathrm{x}}$ [6] under nominal conditions. The set $\mathbb{I}_k$, containing all the information available to the adversary $\mathcal{A}$ at time k, satisfies the following:

Assumption 2 (Strong adversary)**.** *The information of the adversary $\mathcal{A}$ satisfies $\mathbb{I}_k \supseteq \{\mathcal{P}, \mathbb{X}_\mathrm{safe}\}$, for all $k \geqslant 0$.* ◊

The objective of the adversary $\mathcal{A}$ is to design an attack $\boldsymbol{a}_{0:N}$, with $N \in \mathbb{N} \cup \{\infty\}$, satisfying the following:

Definition 1 (Disruptive attack)**.** *An attack $\boldsymbol{a}_{0:N}$ is* disruptive *if there exists a $k' \leqslant N$ for which $\boldsymbol{x}_{k'} \notin \mathbb{X}_\mathrm{safe}$.*

Definition 2 (ϵ-stealthy attack, adapted from [8])**.** *An attack $\boldsymbol{a}_{0:N}$ is* ϵ-stealthy *if, for all $k \leqslant N$, $\|\boldsymbol{C}\boldsymbol{x}_k\| \leqslant \epsilon \in \mathbb{R}_{\geqslant 0}$.*

[1] A multiset is a generalization of a set that allows for duplicate elements, therefore accounting for multiplicity.

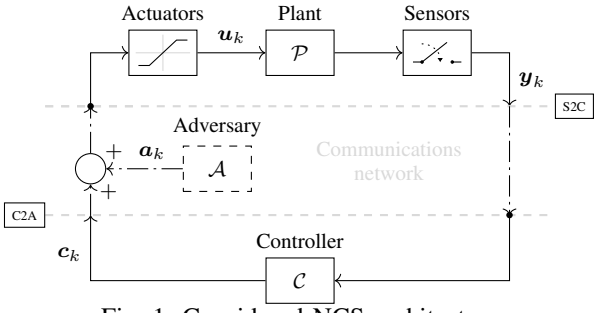


Fig. 1: Considered NCS architecture.

Note that conventional ZDAs are a class of 0-stealthy attacks. If an attack $\boldsymbol{a}_{0:N}$ is both disruptive and stealthy, we call such an attack *successful*, and *unsuccessful* otherwise.

Remark 1. *Conventional ZDAs are 0-stealthy if $\boldsymbol{x}_0 \in \ker(\boldsymbol{C}) \setminus \{\boldsymbol{0}\}$. When $\boldsymbol{x}_0 = \boldsymbol{0}$, whilst no longer 0-stealthy, the transient they induce can be made arbitrary small* [5].

III. Conventional zero dynamics attack

In this section, we consider conventional ZDAs and show that the three common methods of construction lead to identical attack sequences. Here, we do not take saturation into account, implying $\bar{u}_i = -\underline{u}_i = \infty$ for all i, and we ignore the measurement noise for now, implying $\boldsymbol{v}_k = \boldsymbol{0}$ for all k. As ZDAs excite the unstable zeros of the system, we first recall some relevant definitions. For MIMO LTI systems, there exist several notions of zeros, including *transmission zeros*, *invariant zeros*, and *decoupling zeros* [14].

Lemma 1 ([14])**.** *Given a plant $\mathcal{P}$ as in* (1) *satisfying Asm. 1 that is nondegenerate*[2]*, the transmission zeros and invariant zeros are identical and the system has no decoupling zeros.*

Therefore, from Lem. 1, we simply refer to the *zeros* $\mathcal{Z} \subset \mathbb{C}$ of $\mathcal{P}$, which is a multiset with $n_\mathrm{z} = \#(\mathcal{Z}) \leqslant n_\mathrm{x} - 1$.

Various ways of constructing ZDAs have been considered, namely through the use of the Rosenbrock system matrix [6], concepts from geometric control [5] (which allow extensions to non-uniformly sampled-data systems [10]), and thorugh the Byrnes-Isidori normal form [7], [8] (which has also been used to construct ZDAs for nonlinear systems [15] and systems with model uncertainty [8]).

We briefly describe how each method can be used to construct a ZDA $\boldsymbol{a}_{0:\infty}$ and conclude by showing their equivalence. As ZDAs target NMP systems, in order for these attacks to be disruptive, we make the following assumption:

Assumption 3. *The plant $\mathcal{P}$ in* (1) *is* non-minimum phase (NMP)*, meaning* $\max\{|z| \,|\, z \in \mathcal{Z}\} > 1$. ◊

Finally, we make the following technical assumption:

Assumption 4. *The zeros $z \in \mathcal{Z}$ have geometric multiplicity equal to their algebraic multiplicity.* ◊

[2] A system is nondegenerate if it contains a finite number of zeros.

A. Rosenbrock system matrix

Consider the equations given by

$$\underbrace{\begin{bmatrix} \boldsymbol{I}\cdot z - \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{0} \end{bmatrix}}_{\boldsymbol{R}(z)} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{u} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \tag{3}$$

where $\boldsymbol{R}(z)$ is the Rosenbrock system matrix, $z \in \mathbb{C}$, $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{x}}}$, and $\boldsymbol{u} \in \mathbb{C}^{n_{\mathrm{u}}}$. The zeros of $\mathcal{P}$ are those values $z \in \mathcal{Z}$ for which $\boldsymbol{R}(z)$ loses rank. By defining the matrix $\boldsymbol{Z} = \mathrm{diag}(\mathcal{Z})$ and matrices $\boldsymbol{X} \in \mathbb{C}^{n_{\mathrm{x}} \times n_{\mathrm{z}}}$, $\boldsymbol{U} \in \mathbb{C}^{n_{\mathrm{u}} \times n_{\mathrm{z}}}$, such that the columns $\boldsymbol{x} = [\boldsymbol{X}]_{\bullet j}$ and $\boldsymbol{u} = [\boldsymbol{U}]_{\bullet j}$ are the solution to (3) for $[\boldsymbol{Z}]_{jj}$, with $j \in \{1, \dots, n_{\mathrm{z}}\}$, a ZDA $\boldsymbol{a}_{0:\infty}$ can be constructed as follows:

$$\mathcal{A}: \qquad \boldsymbol{a}_k = \boldsymbol{\Phi}_k(\boldsymbol{U}, \boldsymbol{Z})\boldsymbol{\beta}_0, \tag{4a}$$

where $\boldsymbol{\Phi}_k(\boldsymbol{U}, \boldsymbol{Z}) = 1/2 \cdot (\boldsymbol{U}\boldsymbol{Z}^k + \overline{\boldsymbol{U}\boldsymbol{Z}}^k)$, and $\boldsymbol{\beta}_0 = \mathrm{col}(\beta_1, \dots, \beta_{n_{\mathrm{z}}}) \neq \boldsymbol{0}$. Given an initial condition $\boldsymbol{x}_0 = \boldsymbol{\Phi}_0(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$, the solution to (1) with input $\boldsymbol{a}_k$ as in (4a) can be written as

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_k(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0. \tag{4b}$$

B. Controlled invariant subspaces

Another approach makes use of controlled invariant subspaces, the definition of which is given next.

Definition 3 (Controlled invariant subspace [16])**.** *A subspace $V \subseteq \mathbb{R}^{n_{\mathrm{x}}}$ is an $(\boldsymbol{A}, \boldsymbol{B})$*-controlled invariant subspace *if $\boldsymbol{A}V \subseteq V + \mathrm{Im}(\boldsymbol{B})$, or, equivalently, if there exists a matrix $\boldsymbol{F}$, called a* friend *of V, such that $(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F})V \subseteq V$.*

The set of all friend of V is denoted $\mathbb{F}(V)$. The collection of all $(\boldsymbol{A}, \boldsymbol{B})$-controlled invariant subspaces $V \subseteq \ker(\boldsymbol{C})$ admit a unique element of largest dimension $V^\star$. This can then be used to construct a ZDA $\boldsymbol{a}_{0:\infty}$ as follows:

$$\mathcal{A}: \qquad \boldsymbol{f}_{k+1} = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F})\boldsymbol{f}_k, \tag{5a}$$
$$\boldsymbol{a}_k = \boldsymbol{F}\boldsymbol{f}_k, \tag{5b}$$

with $\boldsymbol{F} \in \mathbb{F}(V^\star)$ and initial condition $\boldsymbol{f}_0 \in V^\star \setminus \{\boldsymbol{0}\}$. The zeros of $\mathcal{P}$ are precisely $\mathcal{Z} = \sigma(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F} \,|\, V^\star)$. Note that $\boldsymbol{f}_k \in V^\star$ for all k, and therefore $\boldsymbol{a}_{0:\infty}$ is only influenced by the eigenstructure contained in $V^\star$, which is fixed and cannot be freely assigned by the choice of $\boldsymbol{F} \in \mathbb{F}(V^\star)$ [17].

C. Byrnes-Isidori normal form

Lastly, we discuss the characterization of ZDAs using the Byrnes-Isidori normal form. The system $\mathcal{P}$ in (1) can be written, by means of an invertible similarity transform matrix $\boldsymbol{T} \in \mathbb{R}^{n_{\mathrm{x}} \times n_{\mathrm{x}}}$ [18] such that $\boldsymbol{T}\boldsymbol{x}_k = \mathrm{col}(\boldsymbol{z}_k, \boldsymbol{q}_k)$, as

$$\begin{bmatrix} \boldsymbol{z}_{k+1} \\ \boldsymbol{q}_{k+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S} & \star \\ \boldsymbol{R} & \star \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_k \\ \boldsymbol{q}_k \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{G} \end{bmatrix} (\boldsymbol{c}_k + \boldsymbol{a}_k), \tag{6a}$$

$$\boldsymbol{y}_k = \begin{bmatrix} \boldsymbol{0} & \check{\boldsymbol{C}} \end{bmatrix} \mathrm{col}(\boldsymbol{z}_k, \boldsymbol{q}_k), \tag{6b}$$

with $\check{\boldsymbol{C}} \in \mathbb{R}^{n_{\mathrm{y}} \times n_{\mathrm{x}} - n_{\mathrm{z}}}$ as in [18, 2.2] (dependent on the (vector) relative degree), $\boldsymbol{z}_k \in \mathbb{R}^{n_{\mathrm{z}}}$, $\boldsymbol{q}_k \in \mathbb{R}^{n_{\mathrm{x}} - n_{\mathrm{z}}}$, and where the matrices denoted by "$\star$" are irrelevant for the discussion at hand. A ZDA $\boldsymbol{a}_{0:\infty}$ can then be constructed as [7]:

$$\mathcal{A}: \qquad \boldsymbol{z}_{k+1} = \boldsymbol{S}\boldsymbol{z}_k, \tag{7a}$$
$$\boldsymbol{a}_k = -\boldsymbol{G}^{-1}\boldsymbol{R}\boldsymbol{z}_k, \tag{7b}$$

where $\boldsymbol{z}_k \in \mathbb{R}^{n_{\mathrm{z}}}$. The spectrum $\sigma(\boldsymbol{S})$ are the zeros $\mathcal{Z}$ of $\mathcal{P}$ [19]. We proof equivalence between methods next.

Remark 2. *In order to be disruptive, the adversary must choose $\boldsymbol{\beta}_0$, $\boldsymbol{f}_0$, or $\boldsymbol{z}_0$ such that it excites at least one NMP zero, the existence of which is guaranteed by Asm. 3.*

Proposition 2. *Suppose $\boldsymbol{x}_0 = \boldsymbol{f}_0 = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_0, \boldsymbol{0}) = \boldsymbol{\Phi}_0(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0 \in \ker(\boldsymbol{C}) \setminus \{\boldsymbol{0}\}$. Then, the ZDAs $\boldsymbol{a}_{0:\infty}$ generated by* (4a)*,* (5b)*, and* (7b) *are identical.*

Proof: *Stealthiness:* The proof follows by induction. Suppose $\boldsymbol{x}_k = \boldsymbol{f}_k = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_k, \boldsymbol{0}) = \boldsymbol{\Phi}_k(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0 \in \ker(\boldsymbol{C})$, which implies $\boldsymbol{y}_k = \boldsymbol{0}$ and thus $\boldsymbol{c}_k = \boldsymbol{0}$. We show that $\boldsymbol{x}_{k+1} = \boldsymbol{f}_{k+1} = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{q}_{k+1}, \boldsymbol{0}) = \boldsymbol{\Phi}_{k+1}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$. Starting with (5a), we have that $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{F}\boldsymbol{f}_k = \boldsymbol{f}_{k+1} \in \ker(\boldsymbol{C})$, where the last equality follows from (5b). Next, we consider the attack sequence as in (7b). Given $\mathrm{col}(\boldsymbol{z}_k, \boldsymbol{0}) = \mathrm{col}(\boldsymbol{z}_k, \boldsymbol{q}_k)$, we have $\boldsymbol{q}_{k+1} = \boldsymbol{R}\boldsymbol{z}_k + \boldsymbol{G}(\boldsymbol{0} - \boldsymbol{G}^{-1}\boldsymbol{R}\boldsymbol{z}_k) = \boldsymbol{0}$, and thus $\mathrm{col}(\boldsymbol{z}_{k+1}, \boldsymbol{q}_{k+1}) = \mathrm{col}(\boldsymbol{z}_{k+1}, \boldsymbol{0})$. By definition, $\boldsymbol{x}_{k+1} = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_{k+1}, \boldsymbol{q}_{k+1}) = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_{k+1}, \boldsymbol{0})$. Finally, given that $\boldsymbol{x}_k = \boldsymbol{f}_k = \boldsymbol{\Phi}_k(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$, this means $\boldsymbol{f}_k$ is the linear combination of the vectors contained in $\boldsymbol{X}$ scaled by the zeros $\boldsymbol{Z}$. As $\boldsymbol{Z} = \mathrm{diag}(\sigma(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F} \,|\, V^\star))$, we have that $\boldsymbol{f}_{k+1} = \boldsymbol{x}_{k+1}$ is scaled once more by matrix $\boldsymbol{Z}$, resulting in $\boldsymbol{f}_{k+1} = \boldsymbol{\Phi}_{k+1}(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0 = \boldsymbol{x}_{k+1}$.

Disruptiveness: Note that the spectrum $\sigma(\boldsymbol{A} + \boldsymbol{B}\boldsymbol{F} \,|\, V^\star)$, the spectrum $\sigma(\boldsymbol{S})$, and the diagonal elements of $\boldsymbol{Z}$ are all equal to the zeros $\mathcal{Z}$ of $\mathcal{P}$. As by Asm. 3 at least one of these zeros is unstable, that implies that the dynamics (4b), (7a), and (5a) are unstable, meaning $\lim_{k\to\infty} \|\boldsymbol{f}_k\| = \lim_{k\to\infty} \|\boldsymbol{z}_k\| = \lim_{k\to\infty} \|\boldsymbol{x}_k\| = \infty$.

Equivalence: The induction argument shows that $\boldsymbol{x}_k = \boldsymbol{f}_k = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_k, \boldsymbol{0}) = \boldsymbol{\Phi}_k(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$ implies $\boldsymbol{x}_{k+1} = \boldsymbol{f}_{k+1} = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_{k+1}, \boldsymbol{0}) = \boldsymbol{\Phi}_k(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$, where the base case $\boldsymbol{x}_0 = \boldsymbol{f}_0 = \boldsymbol{T}^{-1}\mathrm{col}(\boldsymbol{z}_0, \boldsymbol{0}) = \boldsymbol{\Phi}_0(\boldsymbol{X}, \boldsymbol{Z})\boldsymbol{\beta}_0$ holds by construction. As the state sequence $\boldsymbol{x}_{0:\infty}$ is identical for all three methods, and the system is assumed to be non-degenerate, the corresponding input sequence $\boldsymbol{a}_{0:\infty}$ generated by (4a), (5b) and (7b) must de identical. ∎

IV. Relaxed zero dynamics attack

The previous section shows how ZDAs, in the absence of saturation, can be successful. To analyze the effects of saturation we consider set of reachable states $\mathbb{X}_{\text{reach}}$ as

$$\mathbb{X}_{\text{reach}} = \left\{ \boldsymbol{x}_k \in \mathbb{R}^{n_{\mathrm{x}}} \;\middle|\; \begin{array}{l} \boldsymbol{x}_0 = \boldsymbol{0}, \boldsymbol{x}_{0:k} \text{ satisfies (1a)}, \\ \underline{u}_i \leqslant [\boldsymbol{u}_k]_i \leqslant \bar{u}_i, \forall k \in \mathbb{N}_0. \end{array} \right\}. \tag{8}$$

Whenever $\underline{u}_i = -\bar{u}_i$ for all $i \in \{1, \dots, n_{\mathrm{u}}\}$, a method to compute an ellipsoidal overapproximation $\bar{\mathbb{X}}_{\text{reach}} = \{\boldsymbol{x} \in \mathbb{R}^{n_{\mathrm{x}}} \,|\, \boldsymbol{x}^\mathsf{T}\boldsymbol{P}\boldsymbol{x} \leqslant n_{\mathrm{y}}\} \supset \mathbb{X}_{\text{reach}}$, where $\boldsymbol{P}$ is the solution to a

[2]Under the restriction that $\mathcal{P}$ has *regular relative degree* [19].

number of linear matrix inequalities (LMIs), is given in [13]. Note that this overapproximation is, in general, tight.

Proposition 3. *Let $\bar{\mathbb{X}}_{\text{reach}} \supseteq \mathbb{X}_{\text{reach}}$ be an overapproximation of the reachable sets under input saturation. If $(\ker(\boldsymbol{C}) \cap \bar{\mathbb{X}}_{\text{reach}}) \subseteq \mathbb{X}_{\text{safe}}$, then conventional ZDAs are unsuccessful.*

Proof: Recall that a conventional ZDA is a 0-stealthy attack, implying $\boldsymbol{x}_k \in \ker(\boldsymbol{C})$ for all k. To be successful, it must hold that $\boldsymbol{x}_k \in \ker(\boldsymbol{C})$ for all $k \leqslant k'$, and that there exists a $k' > 0$ such that $\boldsymbol{x}_{k'} \notin \mathbb{X}_{\text{safe}}$. As $\boldsymbol{x}_k \in \bar{\mathbb{X}}_{\text{reach}}$, whenever $(\ker(\boldsymbol{C}) \cap \bar{\mathbb{X}}_{\text{reach}}) \subseteq \mathbb{X}_{\text{safe}}$, we cannot satisfy $\boldsymbol{x}_{k'} \notin \mathbb{X}_{\text{safe}}$ without violating $\boldsymbol{x}_{k'} \in \ker(\boldsymbol{C})$, making the attack unsuccessful. ∎

The possibility of conventional ZDAs being unsuccessful stems from the fact that, intrinsically, conventional ZDAs are 0-stealthy. This restriction might, however, be overly conservative: in practice, there will always be noise and model uncertainties present and sufficiently small output deviations from nominal might, therefore, not get detected. Therefore, we propose a new type of attack:

Definition 4 (Relaxed zero dynamics attack (ZDA))**.** *Given an $\eta \in \mathbb{R}_{>0}$, a* relaxed ZDA $\boldsymbol{a}_{0:N}$ *satisfies $\boldsymbol{a}_k \neq \boldsymbol{0}$ and $\boldsymbol{x}_k \in \ker(\boldsymbol{C}) \oplus \mathbb{B}_\eta$ for all k, where $\mathbb{B}_\eta = \{\boldsymbol{x} \in \mathbb{R}^{n_\text{x}} \mid \|\boldsymbol{x}\| \leqslant \eta\}$.*

Evidently, relaxed ZDAs are the attacks $\boldsymbol{a}_{0:N}$ which are η-stealthy, as $\boldsymbol{x}_k \in \ker(\boldsymbol{C}) \oplus \mathbb{B}_\eta$ implies $\|\boldsymbol{C}\boldsymbol{x}_k\| \leqslant \eta$ (see Fig. 2). If η is of the same order of the magnitude as the measurement noise $\boldsymbol{v}_k$, the attack might be hard to detect in practice, as it is nearly indistinguishable from the effect of the actual noise [8]. Next, we provide a method to construct relaxed ZDAs by solving a convex optimization problem.

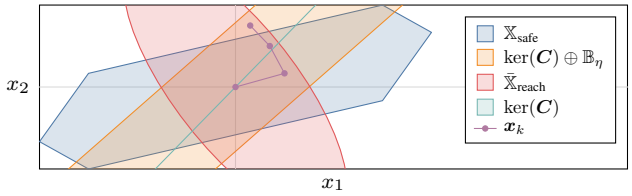


Fig. 2: Conventional 0-stealthy ZDAs are unsuccessful, as $(\ker(\boldsymbol{C}) \cap \bar{\mathbb{X}}_{\text{reach}}) \subseteq \mathbb{X}_{\text{safe}}$, but relaxed ZDAs are successful, as $\boldsymbol{x}_{0:3}$ is entirely contained in both $\ker(\boldsymbol{C}) \oplus \mathbb{B}_\epsilon$ and $\bar{\mathbb{X}}_{\text{reach}}$.

A. Offline optimization

The relaxed ZDA in Definition 4 describes how 0-stealthiness can be relaxed, but does not directly relate to input saturation. To address the latter, we propose to construct a relaxed ZDA by solving the following optimization problem:

$$\min_{\mathbb{D}} f(\mathbb{D}) \quad \text{s.t.} \tag{9a}$$

$$\text{for all } \boldsymbol{x}_0 \in \hat{\mathbb{X}}_0, \text{for all } k \in \{0, \dots, N-1\}: \tag{9b}$$

$$\mathcal{A}: \quad \boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{a}_k, \tag{9c}$$

$$\boldsymbol{a}_k \in \mathbb{U}, \tag{9d}$$

$$\|\boldsymbol{C}\boldsymbol{x}_k\| \leqslant \eta, \tag{9e}$$

$$\boldsymbol{x}_N \notin \mathbb{X}_{\text{safe}}. \tag{9f}$$

In (9a), $\mathbb{D}$ is the set of decision variables, $f : \mathbb{D} \to \mathbb{R}$ is a (convex) cost function, and $\hat{\mathbb{X}}_0 \subset \mathbb{R}^{n_\text{x}}$ is the set of possible initial conditions (which can be a singleton, e.g., $\hat{\mathbb{X}}_0 = \{\boldsymbol{0}\}$). The specification of $\mathbb{D}$ and f depends on the additional objectives of the adversary $\mathcal{A}$ (see Sections IV-B and IV-C), but it always holds that $\{\boldsymbol{a}_{0:N-1}, \boldsymbol{x}_{1:N}\} \subseteq \mathbb{D}$.

With a suitable f, problem (9) can be converted to a (finite number of) second order cone programming (SOCP) problems[3]. Note that $\boldsymbol{x}_N \notin \mathbb{X}_{\text{safe}}$ implies $\boldsymbol{x}_N \in \mathbb{R}^{n_\text{x}} \setminus \mathbb{X}_{\text{safe}}$, which is not a convex set. As $\mathbb{X}_{\text{safe}}$ is assumed to be a polytope (the intersection of m halfspaces), it can be written as

$$\mathbb{X}_{\text{safe}} = \{\boldsymbol{x} \in \mathbb{R}^{n_\text{x}} \mid \text{col}(\boldsymbol{g}_1^\mathsf{T}, \cdots, \boldsymbol{g}_m^\mathsf{T})\boldsymbol{x} \leqslant \boldsymbol{b}\}. \tag{10}$$

By means of a *convex reformulation*, problem (9) is therefore equal to the optimal solution of m convex problems, in which we replace (9f) with $\boldsymbol{g}_i^\mathsf{T}\boldsymbol{x}_N > [\boldsymbol{b}]_i$. The saturation constraint in (2) can easily be converted as $\mathbb{U} = [\underline{u}_1, \bar{u}_1] \times \cdots \times [\underline{u}_{n_\text{u}}, \bar{u}_{n_\text{u}}]$.

Note that (9c) does not take into account the control input $\boldsymbol{c}_k$, which is in general not possible as $\boldsymbol{c}_k \notin \mathbb{I}_k$. This means that the sum $\boldsymbol{a}_k + \boldsymbol{c}_k$ might saturate the actuators, which can lead to detection. In steady-state, usually, one has $\boldsymbol{c}_k \approx \boldsymbol{0}$ and the induced trajectory by the relaxed ZDA remains close to $\ker\boldsymbol{C}$ by design. The adversary $\mathcal{A}$ can introduce additional conservatism by replacing (9d) with $\boldsymbol{a}_k \in \alpha \cdot \mathbb{U}$ and $\alpha \in (0, 1)$. The use of an optimization framework adds versatility to consider auxiliary objectives, of which we highlight two relevant examples next.

B. Shortest and stealthiest attack

Instead of fixing N and η beforehand and checking feasibility of (9), by taking $\eta, N \in \mathbb{D}$ as decision variables and choosing $f(\mathbb{D}) = \text{col}(N, \eta) \in \mathbb{R}^2$, the adversary can find both the stealthiest attack as well as the smallest horizon for which the attack is feasible. Both might be of practical interest, as having a longer attack duration might lead to detection from other sources (i.e., camera or physical surveillance) and less stealthy attacks might lead to incidental detection due to measurement noise.

Remark 3. *If $\mathcal{P}$ has very slow zeros ($\max\{|z| \mid z \in \mathcal{Z}\} \approx 1$), the former gives insight how the adversary can sacrifice stealthiness to decrease the required duration of the attack.*

With cost function $f(\mathbb{D}) = \text{col}(N, \eta)$, problem (9) is a multi-objective optimization problem. In our context, a line search over $N \in \{1, \dots, \bar{N}\}$, with a specified maximum horizon $\bar{N}$, suffices. By keeping N fixed, and minimizing η, the set of feasible solutions forms a Pareto (see Section V).

C. Robustness to various initial conditions

The assumption that $\boldsymbol{x}_0 = \boldsymbol{0}$ is often violated in practice. With a lack of disclosure resources (i.e., $\boldsymbol{x}_0 \notin \mathbb{I}_k$), the adversary can construct a relaxed ZDA which is robust to various initial conditions, by choosing $\hat{\mathbb{X}}_0$ as a polytopic set

[3]SOCPs can be solved efficiently by interior-point methods [20].

for which the adversaries has confidence that $\boldsymbol{x}_0 \in \hat{\mathbb{X}}_0$ [8]. Note that we only need to consider the n_0 vertices $\boldsymbol{x}_{j|0} \in \mathbb{R}^{n_x}$ of $\hat{\mathbb{X}}_0$, and constraints (9c) and (9e) are then given by

$$\boldsymbol{x}_{j|k+1} = \boldsymbol{A}\boldsymbol{x}_{j|k+1} + \boldsymbol{B}\boldsymbol{a}_k, \tag{11a}$$
$$\left\|\boldsymbol{C}\boldsymbol{x}_{j|k}\right\| \leqslant \eta, \tag{11b}$$

respectively, with $\boldsymbol{x}_{j|k} \in \mathbb{D}$ for all $j \in \{1, \dots, n_0\}$. Note that the attack vector $\boldsymbol{a}_{0:N-1}$ is common for all initial conditions, guaranteeing robustness.

V. Illustrative Example

Consider the dynamics of a floating wind turbine (FWT) as given in [21], linearized around the below-rated wind speed $V = 12\,\mathrm{m\,s^{-1}}$. The parameters are taken from the DTU 10 MW reference wind turbine [22]. With a sampling time of $1\,\mathrm{s}$, the discretized dynamics are given by

$$\boldsymbol{A} = \begin{bmatrix} -0.11 & 0.29 & 0.24 \\ -0.62 & 0.54 & 0.82 \\ -0.13 & -0.10 & 0.96 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 0.16 \\ -0.88 \\ 0.19 \end{bmatrix}, \tag{12}$$

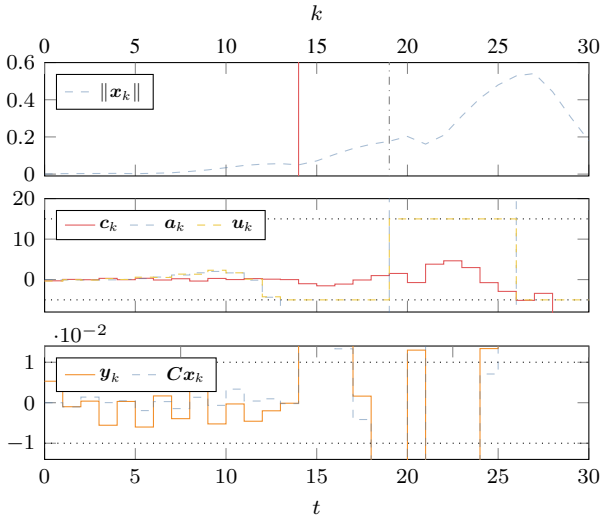
with state $\boldsymbol{x}_k = \mathrm{col}(\Omega_k, \phi_k, \dot{\phi}_k)$, input $\boldsymbol{u}_k = \beta_k$, and output $\boldsymbol{y}_k = \Omega_k$. Here, Ω denotes the generator speed (in RPM), ϕ the platform pitch (in $^\circ$), $\dot{\phi}$ the platform pitch rate (in $^\circ\,\mathrm{s}^{-1}$), and β the collective blade pitch (in $^\circ$).

The discretized system is stable with a pair of complex conjugate zeros at $1.4 \pm 0.64\mathrm{i}$, which are NMP. We define the safe region to be $\mathbb{X}_{\text{safe}} = [-1, 1] \times [-30, 30] \times [-12, 12]$ and the input saturation $\beta \in [-5, 15]$. For the controller $\mathcal{C}$, we take a simple unity feedback control law $\boldsymbol{c}_k = -\boldsymbol{y}_k$. The measurement noise $\boldsymbol{v}_k$ is uniformly distributed on the interval $[-0.01, 0.01]$, and an alarm is raised whenever $\|\boldsymbol{y}_k\| > 0.01$.

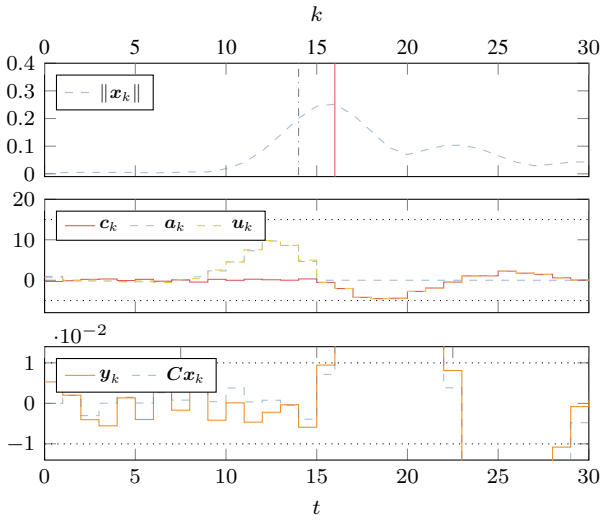
We construct a conventional ZDA using (4a), with $\boldsymbol{\beta}_0 = 10^{-3} \cdot \boldsymbol{1}$. Using (9), we also construct a relaxed ZDA, setting $f(\mathbb{D}) = 0$, $N = 15$, $\hat{\mathbb{X}}_0 = \{\boldsymbol{0}\}$, and $\eta = 0.01$, similar to the measurement noise. The resulting $30\,\mathrm{s}$ simulations with initial condition $\boldsymbol{x}_0 = \boldsymbol{0}$ can be seen in Fig. 3. The red and dashed gray vertical lines depicts the first k for which $\|\boldsymbol{y}_k\| > 0.01$ and the first k for which $\boldsymbol{x}_k \notin \mathbb{X}_{\text{safe}}$, respectively. It is evident that the conventional ZDA is unsuccessful, as the attack is detected before it becomes disruptive. The relaxed ZDA is successful, as it remains stealthy until after the attack has become disruptive.

Next, we consider constructing a relaxed ZDA with cost function $f(\mathbb{D}) = \mathrm{col}(N, \eta)$, leading to a multi-objective optimization problem. The resulting feasibility region from performing a line search over $N \in \{1, \dots, 15\}$ can be seen in Fig. 4. The stealthiest attack is achieved with $\eta \approx 0.0045$, for $N \geqslant 11$. The shortest possible attack sequence has length $N = 5$ but at the cost of larger $\eta \approx 0.042$.

Finally, we consider once more a relaxed ZDA, but now, suppose $\boldsymbol{x}_0 = \mathrm{col}(-0.04, 0.03, -0.01)$, unknown to the adversary. The adversary has some confidence that $\boldsymbol{x}_0 \in \hat{\mathbb{X}}_0 = [-0.05, 0.05]^3$, which will be used to solve (9b). The detection threshold is set to 0.1, as the non-zero initial state will cause a transient. The optimization problem (9) is feasible, and the simulation results can be seen in Fig. 5.



(a) Unsuccessful conventional ZDA



(b) Successful relaxed ZDA

Fig. 3: Simulation result on a FWT.

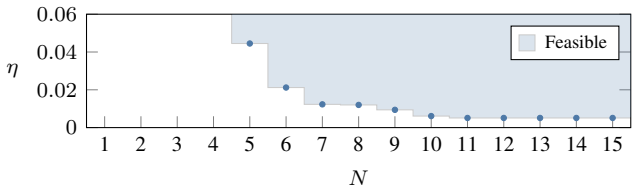


Fig. 4: Pareto front for the FWT with $f(\mathbb{D}) = \mathrm{col}(N, \eta)$.

It is evident that the relaxed ZDA is once again successful, despite inexact knowledge of $\boldsymbol{x}_0$.

VI. Conclusions and Future Work

We have discussed the effect of input saturation on the success of ZDAs. We have shown that, whilst conventional methods for constructing ZDAs can fail under input saturation, by relaxing the 0-stealthy condition, the adversary is still be able to inflict damage to the process whilst remaining practically stealthy. We have proposed an

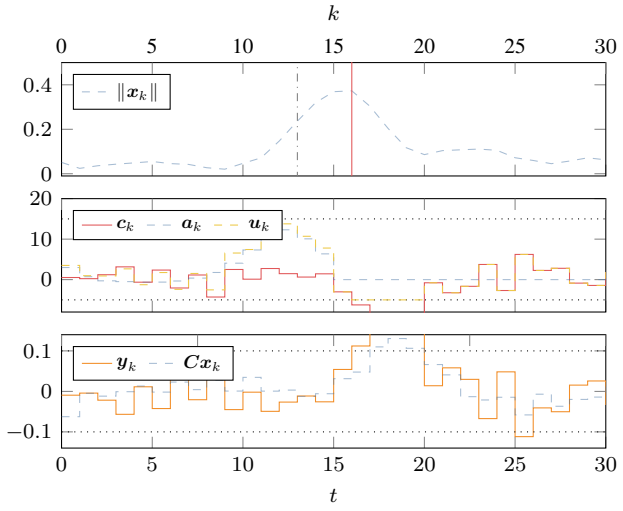


Fig. 5: Simulation result with $\hat{\mathbb{X}}_0 = [-0.05, 0.05]^3$.

optimization framework to construct these relaxed ZDAs, where saturation can be taken directly into account when constructing an attack. We have demonstrated the versatility of the framework and illustrated its effectiveness on a CPSs.

Future work will investigate whether relaxed ZDAs can also be successfully applied to nonlinear systems and systems with model uncertainty. It will also be of interest to further quantify detectability due to non-strict stealthiness, and investigate the vulnerability of systems not in steady state, seeing whether during a transient more states might be reachable, making the relaxed ZDA potentially more dangerous.

REFERENCES

[1] M. Krotofil, "Industrial Control Systems: Engineering Foundations and Cyber-Physical Attack Lifecycle," Information Systems Security Partners (ISSP), Technical white paper, Mar. 2023.

[2] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure Networked Control Systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. Volume 5, 2022, pp. 445–464, May 2022.

[3] C. De Persis and P. Tesi, "Resilient Control under Denial-of-Service," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 134–139, Jan. 2014.

[4] C. Kwon, W. Liu, and I. Hwang, "Security analysis for Cyber-Physical Systems against stealthy deception attacks," in *2013 American Control Conference*. Washington, Washington, USA: IEEE, Jun. 2013, pp. 3344–3349.

[5] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Liverpool, Merseyside, UK: IEEE, Oct. 2012, pp. 1806–1813.

[6] ——, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, Jan. 2015.

[7] J. Kim, J. Back, G. Park, C. Lee, H. Shim, and P. G. Voulgaris, "Neutralizing zero dynamics attack on sampled-data systems via generalized holds," *Automatica*, vol. 113, p. 108778, Mar. 2020.

[8] G. Park, C. Lee, H. Shim, Y. Eun, and K. H. Johansson, "Stealthy Adversaries Against Uncertain Cyber-Physical Systems: Threat of Robust Zero-Dynamics Attack," *IEEE Transactions on Automatic Control*, vol. 64, no. 12, pp. 4907–4919, Dec. 2019.

[9] A. Norouzi Mobarakeh, M. Ataei, and R. Hooshmand, "The threat of zero-dynamics attack on non-linear cyber-physical systems," *IET Cyber-Physical Systems: Theory & Applications*, vol. 1, no. 14, p. cps2.12099, Jun. 2024.

[10] B. Wolleswinkel, M. Mazo, and R. M. G. Ferrari, "Switched Zero Dynamic Attacks on Sampled-Data Systems with Non-uniform Sampling: Vulnerability and Countermeasures," Oct. 2024.

[11] K. Kimura and H. Ishii, "Quantized Zero Dynamics Attacks against Sampled-data Control Systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. Cancún, Quintana Roo, Mexico: IEEE, Dec. 2022, pp. 6140–6145.

[12] M. Naghnaeian, N. H. Hirzallah, and P. G. Voulgaris, "Security via multirate control in cyber–physical systems," *Systems & Control Letters*, vol. 124, pp. 12–18, Feb. 2019.

[13] S. H. Kafash, J. Giraldo, C. Murguia, A. A. Cardenas, and J. Ruths, "Constraining Attacker Capabilities Through Actuator Saturation," in *2018 Annual American Control Conference (ACC)*. Milwaukee, Wisconsin, USA: IEEE, Jun. 2018, pp. 986–991.

[14] J. Tokarzewski, "Zeros in Discrete-Time Mimo Lti Systems and the Output-Zeroing Problem," *International Journal of Applied Mathematics and Computer Science*, no. Vol. 10, no 3, pp. 537–557, 2000.

[15] G. Park, C. Lee, and H. Shim, "On Stealthiness of Zero-Dynamics Attacks against Uncertain Nonlinear Systems," in *23rd International Symposium on Mathematical Theory of Networks and Systems*. Hong Kong, China: Hong Kong University of Science and Technology, Jul. 2018, pp. 10–17.

[16] G. Basile and G. Marro, *Controlled and conditioned invariants in linear system theory*. Englewood Cliffs, NJ: Prentice Hall, 1992.

[17] L. Ntogramatzidis, T. Nguyen, and R. Schmid, "Repeated eigenstructure assignment for controlled invariant subspaces," *European Journal of Control*, vol. 26, pp. 1–11, Nov. 2015.

[18] M. Mueller, "Normal form for linear systems with respect to its vector relative degree," *Linear Algebra and its Applications*, vol. 430, no. 4, pp. 1292–1312, Feb. 2009.

[19] B. Zhou, "On the relative degree and normal forms of linear systems by output transformation with applications to tracking," *Automatica*, vol. 148, p. 110800, Feb. 2023.

[20] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1, pp. 193–228, Nov. 1998.

[21] D. Stockhouse, M. Phadnis, A. Henry, N. Abbas, M. Sinner, M. Pusch, and L. Y. Pao, "Sink or Swim: A Tutorial on the Control of Floating Wind Turbines," in *2023 American Control Conference (ACC)*. San Diego, California, USA: IEEE, May 2023, pp. 2512–2529.

[22] C. Bak, F. Zahle, R. Bitsche, T. Kim, A. Yde, L. Henriksen, A. Natarajan, and M. H. Hansen, "Description of the DTU-10MW reference wind turbine," DTU Wind Energy, Copenhagen, Denmark, Technical Report Tech. Rep. I-0092, 2013, 2013.