This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3318194

Risk Assessment of Stealthy Attacks on Uncertain Control Systems

Sribalaji C. Anand, *Student Member, IEEE*, André M. H. Teixeira, *Member, IEEE*, and Anders Ahlén, *Senior Member, IEEE*

***Abstract*— In this article, we address the problem of risk assessment of stealthy attacks on uncertain control systems. Considering data injection attacks that aim at maximizing impact while remaining undetected, we use the recently proposed output-to-output gain to characterize the risk associated with the impact of attacks under a limited system knowledge attacker. The risk is formulated using a well-established risk metric, namely the maximum expected loss. Under this setup, the risk assessment problem corresponds to an untractable infinite non-convex optimization problem. To address this limitation, we adopt the framework of scenario-based optimization to approximate the infinite non-convex optimization problem by a sampled non-convex optimization problem. Then, based on the framework of dissipative system theory and S-procedure, the sampled non-convex risk assessment problem is formulated as an equivalent convex semi-definite program. Additionally, we derive the necessary and sufficient conditions for the risk to be bounded. Finally, we illustrate the results through numerical simulation of a hydro-turbine power system.**

***Index Terms*— Security, Uncertainty, Risk Analysis, Optimization.**

I. Introduction

Research in the security of industrial control systems has received considerable attention [1] due to an increased number of cyber-attacks such as the one on the Ukrainian power grid [2], Kemuri water company [3] among others. One of the common recommendations for improving the security of control systems is to follow the risk management cycle: risk assessment, risk response, and risk monitoring [4]. This article focuses on risk assessment, the formal definition of which will be introduced later as a function of the attack impact.

Risk is often a combination of attack impact and/or likelihood. For instance, the risk is characterized in terms of average impact in [5] for different types of attacks. The consequences of data injection attacks are quantified using the conditional value-at-risk in [6]. The calculated risk can later be used to compute optimal defense-allocation strategies [7] and/or design robust controllers/detectors [8]. Risk assessment of combined data integrity and availability attacks against the power system state estimation is conducted in [9]. From this brief discussion, it can be realized that characterizing risk in terms of attack impact and likelihood is critical for the efficient allocation of protection resources. In the literature, the problem of risk assessment of stealthy attacks on uncertain control systems has not been addressed. To the best of the author's knowledge, the works that are closely related to this problem are [10], [11], and [12].

Firstly, [10] designs a stealthy attack against an uncertain system using disclosure resources. Secondly, [11] focuses on attack detection based on plant and model mismatch for the adversary. The results of both the above works cannot facilitate the optimal allocation of protection resources.

Thirdly, [12] proposes an impact metric by computing a bound on the reachable set of states by an attacker for perturbed system dynamics. It also proposes a second metric by computing the distance between the reachable set of states for the adversary and the set of critical states. However, it considers a deterministic system.

The advantage of our study is multi-fold. Firstly, we consider a generic modeling framework similar to [13]. Secondly, unlike many previous works, a system description with parametric uncertainty is considered. Thirdly, similar to [7], we adopt an adversarial setup where the system knowledge of the adversary is limited. Finally, we consider a recently proposed impact metric: Output-to-Output Gain (OOG) [14]. The main advantage of using this impact metric, as opposed to the classical H_∞ and $H_{\_}$ metrics, is that the OOG metric-based design problem focuses on improving detectability only when the impact of the attack is sufficiently high at the same frequency [15]. In other words, the OOG metric is more amenable to risk-optimal system design for increased security. Under the described setup, we present the following contributions.

1) Using OOG as an impact metric, and the maximum expected loss as a risk metric, we formulate the risk assessment problem. We observe that the risk assessment problem corresponds to an untractable infinite non-convex robust optimization problem which is NP-hard in general.
2) We propose a convex semi-definite program (SDP) that solves the risk assessment problem approximately by sampling the uncertainty set. Additionally, we provide probabilistic guarantees on the feasibility of the original robust optimization problem.
3) We derive the necessary and sufficient conditions for the risk metric to be bounded.
4) Our approach results in the risk of an open-loop attack which is robust against uncertainties.

The remainder of the article is organized as follows. The uncertain system and the adversary are described in Section II. The problem is formulated in Section III, to which an approximate solution is presented in Section IV. The efficacy of the proposed optimization problem is illustrated through a numerical example in Section V. We conclude the article in Section VI.

Notation: Throughout this article, $\mathbb{R}, \mathbb{R}^+, \mathbb{C}, \mathbb{Z}$ and $\mathbb{Z}^+$ represent the set of real numbers, positive real numbers, complex numbers, integers, and non-negative integers respectively. A positive (semi) definite matrix A is denoted by $A \succ 0 (A \succeq 0)$. Let $x : \mathbb{Z} \to \mathbb{R}^n$ be a discrete-time signal with $x[k]$ as the value of the signal x at the time step k. Let the time horizon be $[0, N] = \{k \in \mathbb{Z}^+ | \ 0 \leq k \leq N\}$. The ℓ_2-norm of x over the horizon $[0, N]$ is represented as $||x||^2_{\ell_2,[0,N]} \triangleq \sum_{k=0}^{N} x[k]^T x[k]$. Let the space of square summable signals be defined as $\ell_2 \triangleq \{x : \mathbb{Z}^+ \to \mathbb{R}^n | \ ||x||^2_{\ell_2,[0,\infty]} < \infty\}$ and the extended signal space be defined as $\ell_{2e} \triangleq \{x : \mathbb{Z}^+ \to \mathbb{R}^n | \ ||x||^2_{[0,N]} < \infty, \forall N \in \mathbb{Z}^+\}$. For the sake of simplicity, we represent $||x||^2_{\ell_2,[0,\infty]}$ as $||x||^2_{\ell_2}$. For $x \in \mathbb{R}$, $\lceil x \rceil$ represents x rounded to the nearest integer greater than or equal to x. Let $(\Omega, \mathcal{D}_a, \mathbf{P})$ represent a probability space with sample space $\Omega \subset \mathbb{R}^v$, $\sigma-$algebra $\mathcal{D}_a$, and probability measure $\mathbf{P}$. Let $(\Omega^w, \mathcal{D}_a^w, \mathbf{P}^w)$ represent the

Manuscript received 31 March 2022; revised 16 August 2023; accepted DD MMMM 2023. This work is supported by the Swedish Research Council under the grant 2018-04396 and by the Swedish Foundation for Strategic Research.

Sribalaji. C. Anand and Anders Ahlén are with the Department of Electrical Engineering, Uppsala University, PO Box 65, SE-75103, Uppsala, Sweden. (email: {sribalaji.anand, anders.ahlen}@angstrom.uu.se)

André M. H. Teixeira is with the Department of Information Technology, Uppsala University, PO Box 337, SE-75105, Uppsala, Sweden. email:andre.teixeira@it.uu.se

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3318194

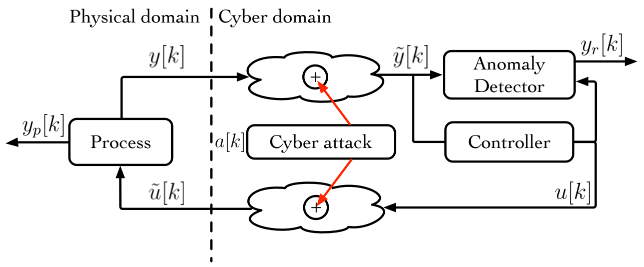

Fig. 1: Control system under data injection attack

w-times Cartesian product of Ω with the $\sigma-$algebra $\mathcal{D}_a^w$ and the probability measure $\mathbf{P}^w = \mathbf{P} \times \cdots \times \mathbf{P}$. A point drawn from $(\Omega^w, \mathcal{D}_a^w, \mathbf{P}^w)$ is thus $(\delta_1, \delta_2, \dots \delta_w)$, i.e., w independent elements in $\mathbb{R}^v$ drawn independently from Ω according to the same probability $\mathbf{P}$[1]. The relative importance of various uncertainty outcomes of an arbitrary function $f(\delta)$, $\delta \triangleq \begin{bmatrix}\delta_1 & \dots & \delta_m\end{bmatrix}^T \in \Omega$, is represented by $\mathbb{P}_\Omega(f(\delta))$. An empty set is denoted by $\emptyset$. For any $a, b \in \mathbb{R}$ and $a \geq b$, $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. The set of non-zero elements of a vector $s \in \mathbb{R}^n$ is denoted by $\text{supp}(s)$. The cardinality of the non-zero elements in a vector $s \in \mathbb{R}^n$ is denoted by $|\text{supp}(s)|$. An indicator function with two expressions is represented by $\mathbb{I}(a, b)$, whose value is 1 when both the expressions are true and zero otherwise.

II. Problem background

In this section, we describe the control system structure and the goal of the adversary. Consider the general description of a closed-loop DT LTI system with a process ($\mathcal{P}$), feedback controller ($\mathcal{C}$), and an anomaly detector ($\mathcal{D}$) as shown in Fig. 1 and represented by

$$\mathcal{P} : \begin{cases} x_p[k+1] &= A^\Delta x_p[k] + B^\Delta \tilde{u}[k] \\ y[k] &= C^\Delta x_p[k] \\ y_p[k] &= C_J x_p[k] + D_J \tilde{u}[k] \end{cases} \tag{1}$$

$$\mathcal{C} : \begin{cases} z[k+1] &= A_c z[k] + B_c \tilde{y}[k] \\ u[k] &= C_c z[k] + D_c \tilde{y}[k] \end{cases} \tag{2}$$

$$\mathcal{D} : \begin{cases} s[k+1] &= A_e s[k] + B_e u[k] + K_e \tilde{y}[k] \\ y_r[k] &= C_e s[k] + D_e u[k] + E_e \tilde{y}[k] \end{cases} \tag{3}$$

where $A^\Delta \triangleq A + \Delta A(\delta)$ with A representing the nominal system matrix and $\delta \in \Omega$. Additionally, we assume Ω to be closed, bounded, and to include the zero uncertainty yielding $\Delta A(0) = 0$. The other matrices are similarly expressed. The state of the process is represented by $x_p[k] \in \mathbb{R}^{n_x}$, $z[k] \in \mathbb{R}^{n_z}$ is the state of the controller, $s[k] \in \mathbb{R}^{n_s}$ is the state of the observer, $\tilde{u}[k] \in \mathbb{R}^{n_u}$ is the control signal received by the process, $u[k] \in \mathbb{R}^{n_u}$ is the control signal generated by the controller, $y[k] \in \mathbb{R}^{n_m}$ is the measurement output produced by the process, $\tilde{y}[k] \in \mathbb{R}^{n_m}$ is the measurement signal received by the controller and the detector, $y_p[k] \in \mathbb{R}^{n_p}$ is the virtual performance output, and $y_r[k] \in \mathbb{R}^{n_r}$ is the residue generated by the detector. In general, the system is considered to have a good performance when the energy of the performance output $||y_p||_{\ell_2}^2$ is small and an anomaly is considered to be detected when the detector output energy $||y_r||_{\ell_2}^2$ is greater than a predefined threshold, say ϵ_r. Without loss of generality (w.l.o.g.), we assume $\epsilon_r = 1$ in the sequel.

A. Data injection attack scenario

In the system described in (1)-(3), we consider an adversary injecting false data into a combination of the sensors and actuators. Next, we discuss the resources the adversary has access to.

[1]In this article, the Cartesian product is considered over the same probability space (Δ). But this can be generalized to arbitrary probability spaces.

1) Disruption and disclosure resources: The adversary can access (eavesdrop) the control or sensor channels and can inject data. This is represented by

$$\begin{bmatrix}\tilde{u}[k] \\ \tilde{y}[k]\end{bmatrix} = \begin{bmatrix}u[k] \\ y[k]\end{bmatrix} + \begin{bmatrix}B_a \\ D_a\end{bmatrix} a[k], \left[\frac{B_a}{D_a}\right] \triangleq \left[\begin{array}{cc} E_a & 0 \\ \hline 0 & F_a \end{array}\right]$$

where $a[k] \in \mathbb{R}^{n_a}$ is the attack signal injected by the adversary. The matrix $E_a(F_a)$ is a diagonal matrix with $E_a(i,i) = 1$ $(F_a(i,i) = 1)$, if the actuator (sensor) channel i is under attack and 0 otherwise.

2) System knowledge: In general, the system operator knows about the parameters of the controller and detector as s/he is the one who designs it. We assume that the adversary can obtain these parameters. But does not know the true parameters of the process. To this end, we consider a realistic adversary whose system knowledge is limited. We defined this adversary as a rational adversary.

Definition 2.1 (Rational adversary): An adversary is defined to be rational if it knows the matrices of (1) with bounded uncertainty. ◁

Defining $x[k] \triangleq [x_p[k]^T \;\; z[k]^T \;\; s[k]^T]^T$, the closed-loop system under attack with the performance output and detection output as system outputs becomes

$$\begin{aligned} x[k+1] &= A_{cl}^\Delta x[k] + B_{cl}^\Delta a[k] \\ y_p[k] &= C_p^\Delta x[k] + D_p^\Delta a[k] \\ y_r[k] &= C_r^\Delta x[k] + D_r^\Delta a[k], \end{aligned} \tag{4}$$

where the nominal matrices are given by $\left[\begin{array}{c|c} A_{cl} & B_{cl} \end{array}\right] \triangleq$

$$\left[\begin{array}{ccc|c} A + BD_cC & BC_c & 0 & BB_a + BD_cD_a \\ B_cC & A_c & 0 & B_cD_a \\ (B_eD_c + K_e)C & B_eC_c & A_e & (B_eD_c + K_e)D_a \end{array}\right]$$

$$C_p \triangleq \begin{bmatrix} C_J + D_JD_cC & D_JC_c & 0 \end{bmatrix}, \quad D_p \triangleq D_J(D_cD_a + B_a),$$
$$C_r \triangleq \begin{bmatrix} (D_eD_c + E_e)C & D_eC_c & C_e \end{bmatrix}, D_r \triangleq (D_eD_c + E_e)D_a.$$

And the uncertain matrices are given as

$$\Delta A_{cl} \triangleq \begin{bmatrix} \Delta A + BD_c\Delta C + \Delta B(D_cC + D_c\Delta C) & \Delta BC_c & 0 \\ B_c\Delta C & 0 & 0 \\ (B_eD_c + K_e)\Delta C & 0 & 0 \end{bmatrix}$$

$$\Delta B_{cl} \triangleq \begin{bmatrix} \Delta BB_a + \Delta BD_cD_a \\ 0 \\ 0 \end{bmatrix}, \quad \Delta C_p \triangleq \begin{bmatrix} D_JD_c\Delta C & 0 & 0 \end{bmatrix}$$

$$\Delta D_p \triangleq 0, \quad \Delta D_r \triangleq 0, \quad \Delta C_r \triangleq \begin{bmatrix} (D_eD_c + E_e)\Delta C & 0 & 0 \end{bmatrix}.$$

Here, the matrices B_a and D_a are chosen by the operator. If all the diagonal elements of B_a or D_a are 1(0), the operator believes that an adversary might attack all (none of the) data channels. Next, we assume the following for clarity.

Assumption 2.1: The closed-loop system (4) is stable $\forall\, \delta \in \Omega$. ◁

Assumption 2.2: The input matrix has full column rank i.e., $\nexists\, s \in \mathbb{R}^{n_a} \neq 0$ such that $B_{cl}^\Delta s = 0$. ◁

3) Attack goals and constraints: Given the resources the adversary has access to, the adversary aims at disrupting the system's behavior while staying stealthy. The system disruption is evaluated by the increase in energy of the performance output. Moreover, similar to the fault detection approaches [16], the adversary is stealthy if the energy of the detection output is below a predefined threshold $\epsilon_r = 1$. Thus we define a stealthy attack as follows.

Definition 2.2 (Stealthy attack): An attack vector a is said to be stealthy if it satisfies $||y_r||_{\ell_2}^2 \leq 1$. ◁

We discuss the attack policy for a deterministic system next.

B. Optimal attack policy for the nominal system

From the previous discussions, it can be understood that the goal of the adversary is to maximize the performance cost while staying undetected. With this setup, we characterize the attack policy of the adversary next before which we introduce the following assumptions.

Assumption 2.3: The closed-loop system (4) is at equilibrium $x[0] = 0$ before the attack commences. $\triangleleft$

Similar to [17], [18], we consider that the adversary has finite energy and thus attacks the system for a long but finite time, say T. Although this time T is unknown *a priori*, it can be ensured that $a[\infty] \triangleq \lim_{k\to\infty} a[k] = 0$ by setting $x[\infty] \triangleq \lim_{k\to\infty} x[k] = 0$.

Lemma 2.1: If $x[\infty] = 0$, then it holds that $a[\infty] = 0$.

Proof: If $x[\infty] = 0$, it follows that $\lim_{k\to\infty} \|x[k+1] - A_{cl}x[k]\| = 0 \implies \lim_{k\to\infty} \|B_{cl}a[k]\| = \|B_{cl}a[\infty]\| = 0$. From *Assumption 2.2*, we known that $B_{cl}a[\infty] = 0$ holds only when $a[\infty] = 0$. This concludes the proof. ∎

In this article, we focus on attacks with finite energy [17], [18] which satisfy $x[\infty] = 0$, and define them as state-vanishing attacks.

Definition 2.3 (State-vanishing attack): An attack vector a is defined to be state-vanishing if applying a to (4) generates the state vector x which satisfies $\lim_{k\to\infty} x[k] = 0$. $\triangleleft$

Then, under *Assumption 2.3*, and the constraint $x[\infty] = 0$ introduced in *Lemma 2.1*, when the system (4) is deterministic, i.e., $\Omega = \emptyset$, [14] formulates the attack policy of the adversary as the following non-convex optimization problem

$$\begin{aligned} ||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r} \triangleq & \sup_{a \in \ell_{2e}} ||y_p||^2_{\ell_2} \\ & \text{s.t.} \quad ||y_r||^2_{\ell_2} \leq 1,\ x[0] = 0,\ x[\infty] = 0, \end{aligned} \tag{5}$$

where $||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r}$ represents the *OOG* that characterizes the disruption caused by the attack signal a. In the literature, such characterization of the consequence of stealthy attacks (5) has only been studied for deterministic systems. The remainder of this article is focused on discussing methods to quantify the consequence of the attack in terms of risk on the uncertain system (4).

The concept of risk is conventionally adopted to address decision-making in an uncertain environment [19]. Since we also focus on decision-making under uncertainty, it is useful to adopt tools from the risk management framework. Thus, before introducing the problem formulation, a brief introduction to risk management and risk metrics is provided as it helps in the problem formulation.

III. PROBLEM FORMULATION

The framework of risk management can help the system operator answer the following (but not limited) questions: (i) Which components of the system are critical to the operation of the system? (ii) What disruption can be expected from attacks and (iii) Which resources should be protected and how? Thus, to use the risk management framework for the benefit of the system operator (to estimate system disruption (ii)), we will focus on risk and its consequences. To quantify the risks of data injection attacks on an uncertain system, we start by defining an impact random variable as a function of the system uncertainty and the attack vector.

Definition 3.1 (Impact random variable): Let the random variable $X^A(\cdot,\cdot)$ be defined as

$$X^A(a,\delta) \triangleq \|y_p(\delta)\|^2_{\ell_2} \times \mathbb{I}\Big(\|y_r(\delta)\|^2_{\ell_2} \leq 1, x(\delta)[\infty] = 0\Big) \tag{6}$$

where $X^A(\cdot,\cdot)$ is the impact caused on the system (4) with uncertainty $\delta \in \Omega$ by the attack $a \in \ell_{2e}$, $\mathbb{I}$ is the indicator function, $y_p(\delta), y_r(\delta)$ and $x(\delta)$ are the performance, residue output, and state of the system with the isolated uncertainty δ. Here, the signals $y_p(\delta), y_r(\delta)$ and $x(\delta)$ are also functions of the attack vector a. $\triangleleft$

With the random variable defined in *Definition 3.1*, we next formulate the risk assessment problem for a rational adversary. Consider the data injection attack scenario where only the bounds of the parametric uncertainty set Ω are known to the adversary. Then the adversary can determine the attack vector which maximizes the expected loss over the entire uncertainty set Ω i.e., the adversary maximizes the function[2]

$$\mathbb{E}_\Omega\left\{X^A(a,\delta)\right\}.$$

This setup is common in game-theoretic approaches [20] where the players do not know the strategy of the other players and thus play by maximizing its expected return over all the strategies of the other players. Similarly, since the adversary has limited system knowledge, s/he chooses an attack policy that maximizes the expected loss of the system operator over the set Ω whilst remaining stealthy. This strategy of maximum expected loss can be defined as follows.

Definition 3.2 (Risk Metric: Maximum Expected Loss ($\mathbb{MEL}$)): The maximum expected loss associated with the impact-random variable $X^A(\cdot)$, defined in (6), is defined as

$$\mathbb{MEL}[X^A] \triangleq \sup_{a \in \ell_{2e}} \mathbb{E}_\Omega\left\{X^A(a,\delta)\right\},$$

where $X^A(\delta, a)$ is the loss on scenario δ and $\mathbb{E}_\Omega$ represents the expectation operator over the set Ω. $\triangleleft$

Thus by determining the attack vector that solves for maximal expected loss, one can ensure that the attack vector is stealthy with respect to all uncertainties whilst maximizing the performance loss. Using *Definition 3.2*, the risk associated with the impact caused by a bounded-**R**ational **A**dversary can be characterized as

$$\gamma_{RA} \triangleq \sup_{a \in \ell_{2e}} \mathbb{E}_\Omega(X^A(a,\delta)). \tag{7}$$

Since the operator $\mathbb{E}$ in (7) operate over the continuous space Ω, (7) is computationally intensive or in general NP-hard [21, Section 3]. Besides, the problem is also non-convex. In the remainder of this article, we discuss methods to solve the optimization problem approximately and efficiently.

IV. RISK ASSESSMENT FOR A BOUNDED-RATIONAL ADVERSARY

In this section, we focus on describing a scenario-based approach to the optimization problem (7).

A. Approximating the uncertainty set

To recall, we are interested in determining the maximum expected loss associated with the impact caused by a rational adversary. Unfortunately, this problem is computationally intensive or in general NP-hard. Thus, as a first step toward solving (7), we approximate the objective function in *Lemma 4.1*.

Lemma 4.1: Let δ_i be sampled uncertainties from Ω. Let us define

$$\hat{\mathbb{E}}_{\Omega_{N_2}}(X^A(a,\delta)) \triangleq \frac{1}{N_2}\sum_{i=1}^{N_2} X^A(a_i,\delta_i).$$

Then, it holds that $\lim_{N_2\to\infty} \hat{\mathbb{E}}_{\Omega_{N_2}}(X^A(a_i,\delta_i)) = \mathbb{E}_\Omega(X^A(a,\delta))$.

Proof: The proof follows from applying [22, Theorem 7.2] to approximate the expectation operator in (7). ∎

[2]In general, the knowledge of the probability distribution of Ω is necessary to determine the expectation. However, as we show in *Theorem 4.3*, the results of this article are independent of the distribution of Ω.

Lemma 4.1 states that the continuous set Ω can be approximated by a discrete set Ω_{N_2} of cardinality N_2. The approximation becomes more accurate as $N_2 \to \infty$. For most cases, this approximation introduces a curse of dimensionality to obtain a good estimate of the risk and obtain a well-feasible attack vector. To circumvent this practical issue, we next show that an attack vector obtained by solving (7) with a discrete uncertainty set as mentioned in *Lemma 4.1* is partially feasible to the original optimization problem (7) with a continuous uncertainty set.

It might not be immediately apparent that the notion of feasibility applies to (7) since there are no external constraints present. Thus, we begin by simplifying the optimization problem (7).

Lemma 4.2: The optimization problem (7) is equivalent to (8).

$$\begin{aligned} \sup_{a\in\ell_{2e},\beta\in[0,\,1]} \quad & \mathbb{E}_\Omega\left\{||y_p(\delta)||^2_{\ell_2} \mid \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\right\}(1-\beta) \\ \text{s.t.} \quad & \mathbb{P}_\Omega\begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\geq 1-\beta. \end{aligned} \tag{8}$$

Proof: Consider the function $X^A(a,\delta)$ in (6). By expanding its indicator function, we can write $\mathbb{E}_\Omega(X^A(a,\delta))$ as

$$\mathbb{E}_\Omega\left\{||y_p(\delta)||^2_{\ell_2} \mid \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\right\}\mathbb{P}_\Omega\begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}$$

Using the above definition in (7), we obtain γ_{RA} as

$$\sup_{a\in\ell_{2e}} \mathbb{E}_\Omega\left\{||y_p(\delta)||^2_{\ell_2} \mid \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\right\}\mathbb{P}_\Omega\begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}$$

which can be rewritten as (8). This concludes the proof. ■

Lemma 4.2 uncovers the constraints present in the optimization problem (7). We can now discuss the notion of feasibility in regard to the optimization problem (8). So, we continue by simplifying (8) as follows. In reality, β represents the fraction of the uncertainty set with respect to which the adversary is not stealthy. Let us consider an adversary that wishes to be stealthy w.r.t. all bounded uncertainties. Thus we could set $\beta=0$. Motivated by the above argument, we rewrite (8) as

$$\begin{aligned} \sup_{a\in\ell_{2e}} \quad & \mathbb{E}_\Omega\left\{||y_p(\delta)||^2_{\ell_2} \mid \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\right\} \\ \text{s.t.} \quad & \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\forall\delta\in\Omega \end{aligned} \tag{9}$$

Recalling the approximation result in *Lemma 4.1*, assume that Ω is replaced with a discrete uncertainty set Ω_N, so that (9) is approximated by (10) whose value is denoted by $\gamma_{RA}(N_2)$.

$$\sup_{a\in\ell_{2e}}\left\{\hat{\mathbb{E}}_{\Omega_{N_2}}\left[||y_p(\delta)||^2_{\ell_2}\right] \Big| \begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\forall\delta\in\Omega_{N_2}.\right\} \tag{10}$$

Let the resulting optimal attack vector be denoted by $a^*_{N_2}$. Then the following theorem provides *a posteriori* results on the feasibility of the attack vector $a^*_{N_2}$ to (9).

Theorem 4.3: Let the number of samples N_2 and the confidence level $\lambda\in(0,1)$ be predefined constants. Define $\epsilon(\cdot)$ such that

$$\epsilon(N_2)=1\;,\;\sum_{k=0}^{N_2-1}\binom{N_2}{k}(1-\epsilon(k))^{N_2-k}=\lambda. \tag{11}$$

Let $s^*_{N_2}$ represent the cardinality of the support subsample for $(\delta_1,\dots,\delta_{N_2})$ (see [23, *Definition 2*]). Then it holds that

$$\mathbb{P}^{N_2}\{\mathbb{P}_\Omega\{\delta\in\Omega\mid a^*_{N_2}\notin\Theta\}>\epsilon(s^*_{N_2})\}\leq\lambda,$$

TABLE I: Dimension of matrices, $n_c \triangleq n_x+n_z+n_s$

Matrix	Dimension	Matrix	Dimension
$\bar{A}$	$n_cN_2\times n_cN_2$	$\bar{B}$	$n_cN_2\times n_a$
$\bar{C}_p$	$n_pN_2\times n_cN_2$	$\bar{D}_p$	$n_pN_2\times n_a$
$\bar{C}_r$	$n_rN_2\times n_cN_2$	$\bar{D}_r$	$n_rN_2\times n_a$

where $a^*_{N_2}$ is the argument of the optimization problem (10) and Θ is defined as

$$\Theta\triangleq\bigcap_{\delta\in\Omega}\Theta_\delta,\text{ where }\Theta_\delta\triangleq\left\{a\in\ell_{2e}\Big|\begin{pmatrix}||y_r(\delta)||^2_{\ell_2}\leq 1\\ x(\delta)[\infty]=0\end{pmatrix}\right\} \tag{12}$$

Proof: See Appendix A.1. ■

In words, *Theorem 4.3* states that the attack vector obtained by solving (10) is stealthy and state-vanishing for all the closed-loop system of the form (4) with uncertainties belonging to the set Ω except for the fraction $\epsilon(s^*_{N_2})$ of Ω. It also states that the $\epsilon(\cdot)$ and λ are independent of the distribution of Ω. This result is a direct consequence of [23, *Theorem 1*].

In conclusion, it follows that (9) can be solved approximately with a discrete set Ω_{N_2} of arbitrary but bounded cardinality. Thus, the next section will focus on solving the optimization problem (10).

B. Risk assessment

The optimization problem (10) has two main disadvantages namely (i) it is a non-convex optimization problem, and (ii) its constraints lie in the infinite-dimensional attack space. To resolve these disadvantages, we adopt the S-procedure and dissipative system theory and revisit the optimization problem (10) in the theorem below. To begin with, given a sampled uncertainty $\delta_i\in\Omega$, we define $\tilde{\Sigma}_{p,i}\triangleq(A_{cl,i},B_{cl,i},C_{p,i},D_{p,i})$ and $\tilde{\Sigma}_{r,i}\triangleq(A_{cl,i},B_{cl,i},C_{r,i},D_{r,i})$ with $y_p(\delta_i)=y_{pi},y_r(\delta_i)=y_{ri}$ and $x(\delta_i)=x_i$ as the outputs and states of $\tilde{\Sigma}_{p,i}$ and $\tilde{\Sigma}_{r,i}$ correspondingly. Now we are ready to present the main theorem of this section.

Theorem 4.4: Let N_2 be a predefined constant. The maximum expected loss (10) caused by a rational adversary injecting a maximally robust attack on (4) can be obtained by solving the convex SDP [3]

$$\begin{aligned} \min_{\gamma\geq 0,P=P^T} \quad & \mathbf{1}^T\begin{bmatrix}\gamma_1 & \dots & \gamma_{N_2}\end{bmatrix}^T \\ \text{s.t.} \quad & \begin{bmatrix}\bar{A}^TP\bar{A}-P & \bar{A}^TP\bar{B}\\ \bar{B}^TP\bar{A} & \bar{B}^TP\bar{B}\end{bmatrix}+\Psi(\gamma)\preceq 0, \end{aligned} \tag{13}$$

where $\Psi(\gamma)\triangleq\frac{1}{N_2}\begin{bmatrix}\bar{C}_p^T\\ \bar{D}_p^T\end{bmatrix}\begin{bmatrix}\bar{C}_p & \bar{D}_p\end{bmatrix}-\begin{bmatrix}\bar{C}_r^T\\ \bar{D}_r^T\end{bmatrix}\Gamma(\gamma)\begin{bmatrix}\bar{C}_r & \bar{D}_r\end{bmatrix}$,

$$\left[\begin{array}{c|c}\bar{A} & \bar{B}\\ \hline \bar{C}_p & \bar{D}_p\\ \hline \bar{C}_r & \bar{D}_r\end{array}\right]=\left[\begin{array}{ccc|c} A_{cl,1} & \dots & 0 & B_{cl,1}\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & A_{cl,N_2} & B_{cl,N_2}\\ \hline C_{p,1} & \dots & 0 & D^T_{p,1}\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & C_{p,N_2} & D^T_{p,N_2}\\ \hline C_{r,1} & \dots & 0 & D^T_{r,1}\\ \vdots & \ddots & \vdots & \vdots\\ 0 & \dots & C_{r,N_2} & D^T_{r,N_2}\end{array}\right],$$

and $\Gamma(\gamma)=I_{n_r}\bigotimes\text{diag}(\gamma_1,\dots\gamma_{N_2})$. The dimension of each matrix is given in TABLE I.

[3]With abuse of notation, we denote that every element of the vector γ is non-negative by $\gamma\geq 0$

Proof: See Appendix A.2. ■

The optimization problem (10) is the primal problem with its optimizer being the attack vector a. This optimization problem is non-convex. With the help of the S-procedure and dissipative system theory, (10) is converted to its equivalent dual SDP form (13) with its optimizer γ, P, which is convex. This equivalency also helps us to conclude that the duality gap is zero. The necessary and sufficient conditions for the value of (13) to be finite is given *Lemma 4.5*.

Lemma 4.5 (Boundedness): Consider N_2 i.i.d. realizations of the closed-loop system (4), each with an uncertainty δ_i. The optimal value of (13) with the above-mentioned system realizations is bounded if and only if the system with $\bar{\Sigma}_p = (\bar{A}, \bar{B}, \bar{C}_p, \bar{D}_p)$ and $\bar{\Sigma}_r = (\bar{A}, \bar{B}, \bar{C}_r, \bar{D}_r)$ satisfy one of the following:

1) The system $\bar{\Sigma}_r$ has no zeros on the unit circle.
2) The zeros on the unit circle of the system $\bar{\Sigma}_r$ (including multiplicity and input direction) are also zeros of $\bar{\Sigma}_p$.

Proof: See Appendix A.3. ■

Let the outputs of $\bar{\Sigma}_p$ and $\bar{\Sigma}_r$ be represented by $\bar{y}_r$ and $\bar{y}_p$ respectively. Then, in words, *Lemma 4.5* states that the maximum expected loss (10) is bounded if either, there does not exist an attack vector which makes the output $\bar{y}_r$ identically zero, or for all attack vectors which yield $\bar{y}_r$ identically 0, it also yields $\bar{y}_p$ identically zero.

It is important to study the conditions for unbounded risk because, (a) if the conditions of the lemma do not hold, it means that there exists an attack vector that can remain stealthy but cause very huge system disruptions, and (b) it depicts that the conditions for unbounded risk are restrictive in comparison to the deterministic case [14, Theorem 2]. However, as the conditions of the lemma are necessary and sufficient, the operator can alter the system matrices so that the conditions hold and consequently reduce the vulnerability of the system to such attacks. In the next section, we illustrate the results with a numerical example.

V. NUMERICAL EXAMPLE

Consider a power generating system [10, Section 4] as shown in Fig. 2 and represented by

$$\begin{bmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \\ \dot{\eta}_3 \end{bmatrix} = \begin{bmatrix} \frac{-1}{T_{lm}} & \frac{K_{lm}}{T_{lm}} & \frac{-2K_{lm}}{T_{lm}} \\ 0 & \frac{-2}{T_h} & \frac{6}{T_h} \\ \frac{-1}{T_g R} & 0 & \frac{-1}{T_g} \end{bmatrix} \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}}_{\eta} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{T_g} \end{bmatrix} \tilde{u} \tag{14}$$

$$y = \underbrace{\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}}_{C} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}, \quad y_p = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{C_p} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}. \tag{15}$$

Here $\eta \triangleq [df; dp + 2dx; dx]$, df is the frequency deviation in Hz, dp is the change in the generator output per unit (p.u.), and dx is the change in the valve position p.u.. The parameters of the plant are listed in TABLE II. The constants T_{lm}, T_h, and T_g represent the time constants of load and machine, hydro turbine, and governor, respectively (See Fig. 2), and $R(Hz/p.u.)$ is the speed regulation due to the governor action. The constant K_{lm} represents the steady-state gain of the load and machine. The DT system matrices $(A^{\Delta}, B^{\Delta}, C^{\Delta}, D^{\Delta})$ are obtained by discretizing the plant (14)-(15) using zero-order hold with a sampling time T_s. The plant is stabilized with an output feedback controller of the form (2) with $D_c = 19$. The detector is an observer-based residue generator of the form (3) with matrices $A_e = (A_d - K_e C_d), B_e = B_d, C_e = C_d$ where $K_e = \begin{bmatrix} 0.17 & -2.83 & -7.43 \end{bmatrix}^T$. In this particular setup, the adversary is considered to attack only the actuator, i.e., $B_a = 1$ and $D_a = 0$. The other unspecified matrices are zero. Only the matrix A^{Δ} is a function of the variable T_h and thus is uncertain. Next, considering (14)-(15) where T_h is uncertain and is uniformly distributed, we discuss the risk of the operator.

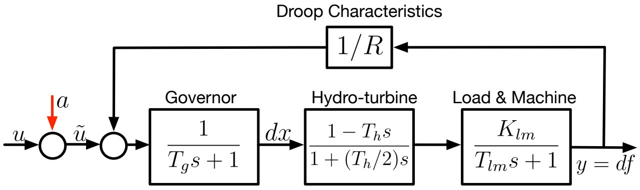

Fig. 2: Power generating system with a hydro turbine.

TABLE II: System Parameters

K_{lm}	1	T_{lm}	6	T_g	0.2
T_h	[4 6]	T_s	0.1	R	0.05

In view of *Lemma 4.1*, we choose $N_2 = 21$ to approximate the set Ω. With this approximation, by solving the convex SDP (13), we obtain $\gamma_{RA}(N_2) = 617.267$. To recall, γ_{RA} represents the maximum expected performance loss of the system operator. In this implementation, the value of risk was obtained for $\beta = 0$ since we considered a maximally robust adversary.

Next, we discuss the validity of the approximation Ω_{N_2}. For varying values of N_2, the number of non-zero γ_is obtained while solving (13) is shown in the first two columns of TABLE III. In view of *Theorem 4.3*, if we solve the problem (13) with an arbitrary N_2, we can provide guarantees on the optimization problem (7) as shown in column 3 of TABLE III. Here, the function $\epsilon(\cdot)$ is evaluated according to [23, (7)] where $\lambda = 10^{-2}$ and $s^*_{N_2} = |\text{supp}(\gamma)|$[4]. And $\epsilon(\cdot)$ should be interpreted as follows: The attack vector obtained by solving (13) with N_2 samples, with probability $1 - \lambda$, will be at most feasible for the fraction $(1 - \epsilon(\cdot))\Omega$ of the set Ω.

Since $\epsilon(\cdot)$ represents the fraction of the uncertainty set to which the attack vector is infeasible, it is intuitive to expect this value to be close to zero. Numerically we have observed from TABLE III that supp(γ) is always one. So, assuming again that $s^*_{N_2} = 1$, the number of samples required to guarantee that the attack vector, with a probability $1 - \lambda$, is feasible for $1 - \epsilon(\cdot)$ of the uncertainty set can be obtained by picking N_2 such that (16) holds [23, (7)]

$$\sqrt[N_2 - 1]{\frac{\lambda}{N_2^2}} = 1 - \epsilon(\cdot). \tag{16}$$

Thus (16) gives an idea of how N_2 increases as $\epsilon(\cdot)$ decreases. And consequently gives an idea of the scalability of the proposed approach as the dimension of the matrix inequality (13) depends on N_2. We also depict the computation complexity by providing the time taken to solve (13) in the last column of TABLE III.

VI. CONCLUSION AND FUTURE WORK

In this article, we study the problem of risk assessment on uncertain control systems under a bounded-rational adversary. Using the OOG

[4]The confidence is denoted by λ here whereas it is denoted by β in [23]

TABLE III: Rational adversary - a posteriori Guarantees

N_2	$\lvert\text{supp}(\gamma)\rvert$	$\epsilon(s^*_{N_2})$	$\gamma_{RA}(N_2)$	Time (seconds)
8	1	0.7141	638.04	17
15	1	0.5112	628.60	116
21	1	0.4142	617.26	578

as an impact metric, we formulated the risk assessment problem and observe that it corresponds to a non-convex robust optimization problem. A scenario-based approach was used to relax the robust optimization problem to their sampled counterpart. Using dissipative system theory, the non-convex sampled problem in signal space was converted to its convex dual problem in matrix inequalities. Detailed proof of the zero-duality gap was provided using the S-procedure. We additionally provide necessary and sufficient conditions for risks to be bounded which highlights the important role of uncertainty and how it is incorporated in attack scenarios. The results are depicted with numerical examples. Future work includes (i) investigating the risk assessment problem where the uncertainty set can be approximated as a polytopic set, and (ii) studying the relation between our risk measure 7 and a coherent risk measure [24].

APPENDIX

A.1. PROOF OF *Theorem 4.3*

Before presenting the proof, an introduction to scenario-based constraint satisfaction is provided.

Scenario-based constraint satisfaction [23]

Consider the constrained non-convex optimization problem $\inf_{\theta\in\Theta} f(\theta,\delta)$, where $\delta \in \Delta$ is the uncertainty and θ is the infinite-dimensional decision parameter which lies in the set $\Theta \triangleq \bigcap_{\delta\in\Delta}\Theta_\delta$, where Θ_δ is the constraint set which includes all the admissible parameters for the isolated uncertainty δ.

Definition A.1.1 (Violation probability): Let us define the violation probability as $\mathbb{V}(\theta) \triangleq \mathbb{P}_\Delta\{\delta \in \Delta \mid \theta \notin \Theta\}$. ◁

Definition A.1.2 (ϵ-level solution): Let $\epsilon \in (0,1)$. Then, $\theta \in \Theta$ is an ϵ level solution if $\mathbb{V}(\theta) \leq \epsilon$. ◁

Definition A.1.3 (Confidence level : λ): Let $\lambda \in (0,1)$. Then, the confidence level λ represents the probability that θ is not an ϵ level solution. i.e., $\lambda \triangleq \mathbb{P}\{\mathbb{V}(\theta) > \epsilon\}$. ◁

Proof: [Proof of *Theorem 4.3*] The optimization problem (9) can be reformulated as

$$-\inf_{a\in\ell_{2e}} \left\{\mathbb{E}_\Omega\{-||y_p(\delta)||^2_{\ell_2}\} \mid \begin{pmatrix} ||y_r(\delta)||^2_{\ell_2} \leq 1 \\ x(\delta)[\infty] = 0\end{pmatrix} \forall \delta \in \Omega\right\} \quad (17)$$

In view of [23, Theorem 1], let us define the objective function as $f(a,\delta) \triangleq \mathbb{E}_\Omega\{-||y_p(\delta)||^2_{\ell_2}\}$, where $y_p(\delta)$ is also a function of the attack vector a. Let us define the set Θ as in (12). Let us define a confidence level $\lambda \in (0,1)$, a constant N_2 and $\epsilon(\cdot)$ such that (11) holds. Then applying [23, Theorem 1], we obtain that

$$\mathbb{P}^{N_2}\{\mathbb{P}_\Omega\{\delta \in \Omega \mid a^*_{N_2} \notin \Theta\} > \epsilon(s^*_{N_2})\} \leq \lambda,$$

Thus, with a probability level $1-\lambda$, the solution $a^*_{N_2}$ is $\epsilon(s^*_{N_2})$ feasible to the optimization problem (17). In our problem setting, $a^*_{N_2}$ is the optimal argument of the optimization problem

$$-arg \inf_{a\in\ell_{2e}} \left\{ \frac{-1}{N_2}\sum_{i=1}^{N_2}\{||y_{p,i}||^2_{\ell_2}\} \middle| \begin{matrix} ||y_{r,i}||^2_{\ell_2} \leq 1 \\ x_i[\infty] = 0 \end{matrix}, \forall i \in \mathcal{S} \right\}$$

where $\mathcal{S} \triangleq \{1,\dots,N_2\}$ which can be rewritten as (10). This concludes the proof. ■

A.2. PROOF OF *Theorem 4.4*

A core step of the proof relies on [25, Theorem 4.3.1], which in turn leverages the notion of S-system. Therefore, before presenting the proof, an introduction to the S-system is provided. Readers interested in a more detailed proof are referred to the extended preprint of this article [26, Section A.5].

S-system [25, Definition 4.3.1]

Let $\mathcal{L}$ be a real Hilbert space with a well-defined inner product denoted by $\langle\cdot,\cdot\rangle$. Let $\mathcal{G}_0(\cdot),\dots,\mathcal{G}_k(\cdot)$ be quadratic functionals mapping $\mathcal{L} \to \mathbb{R}$. Let ω be a discrete-time signal.

Definition A.2.1 (S-system): [25, *Definition 4.3.1*] The quadratic functionals $\mathcal{G}_0(\cdot),\mathcal{G}_1(\cdot),\dots,\mathcal{G}_k(\cdot)$ form an S-system if there exist a bounded linear operator $\mathbf{T}_i : \mathcal{L} \to \mathcal{L}, i = 1,2,\dots,$ such that

1) $\langle \mathbf{T}_i\omega_1, \omega_2\rangle \to 0$ as $i \to \infty$, $\forall\, \omega_1,\omega_2 \in \mathcal{L}$.
2) If $\omega \in \mathcal{M}$, then $\mathbf{T}_i\omega \in \mathcal{M}$, $\forall i = 1,2,\dots,$ where $\mathcal{M}$ is a linear subspace of $\mathcal{L}$.
3) $\mathcal{G}_j(\mathbf{T}_i\omega) \to \mathcal{G}_j(\omega)$ as $i \to \infty$, $\forall \omega \in \mathcal{L}$ and $j = 0,1,\dots,k$. ◁

We next present a theorem that helps in proving *Theorem 4.4*.

Theorem A.2.1: Let us define a stable discrete-time linear time-invariant system of the form

$$\begin{aligned} \eta[k+1] &= \Phi\eta[k] + \Lambda\mu[k] \\ \sigma[k] &= \Pi\eta[k] + \Upsilon\mu[k] \qquad \eta[0] = \eta_0,\ \eta[\infty] = 0. \end{aligned} \quad (18)$$

Let us define the set $\mathcal{L}$ as

$$\mathcal{L} \triangleq \left\{\omega = \begin{bmatrix}\sigma \\ \mu\end{bmatrix} \mid \begin{bmatrix} \sigma,\mu \text{ are related by (18)} \\ \mu \in \ell_{2e},\ \eta[0] = \eta_0,\ \eta[\infty] = 0. \end{bmatrix}\right\} \quad (19)$$

Let us also define the functionals $\mathcal{G}_0(\omega) \triangleq \sum_{k=0}^{\infty} \omega[k]^T M_0 \omega[k] + \zeta_0,\dots,\mathcal{G}_k(\omega) \triangleq \sum_{k=0}^{\infty} \omega[k]^T M_k\omega[k] + \zeta_k$. where $M_0,\dots,M_k$ are given matrices and $\zeta_0,\dots,\zeta_k$ are given constants. Then, the functionals $-\mathcal{G}_0(\cdot),\dots,\mathcal{G}_k(\cdot)$ form an S-system.

Proof: [Proof of *Theorem A.2.1*] In view of *Definition A.2.1*, let us define the operator $\mathbf{T}_i$ as

$$\mathbf{T}_i\omega[k] = \begin{cases} 0, & \text{if } 0 \leq k \leq i \\ \omega[k-i], & \text{if } k > i \end{cases}.$$

For any $\omega_1,\omega_2 \in \mathcal{L}$, $\langle \mathbf{T}_i\omega_1,\omega_2\rangle = |\sum_{k=0}^{\infty}\langle \mathbf{T}_i\omega_1[k],\omega_2[k]\rangle|^2 =$

$$|\sum_{k=i}^{\infty}\langle\omega_1[k-i],\omega_2[k]\rangle|^2 \leq_1 \sum_{k=i}^{\infty}||\omega_1[k-i]||^2\sum_{k=i}^{\infty}||\omega_2[k]||^2$$

where the inequality 1 holds due to the Cauchy-Schwartz inequality. Since the theorem states that $\lim_{k\to\infty}\eta[k] = 0$, it holds from *Lemma 2.1* that $\lim_{k\to\infty}\mu[k] = 0$. Following which, it immediately holds that $\lim_{k\to\infty}\sigma[k] = 0$, which implies that $\lim_{k\to\infty}\omega[k] = 0$. Due to the above reasoning, it holds that $\lim_{i\to\infty}\sum_{k=i}^{\infty}||\omega_2[k]||^2 = 0\ \forall \omega \in \mathcal{L}$. Thus condition 1) of *Definition A.2.1* holds.

Let us consider a set $\mathcal{M} = \mathcal{L}|_{\eta_0=0}$. Then, if $\omega \in \mathcal{M}$, due to the time-invariant property of (18), $\mathbf{T}_i\omega \in \mathcal{M}$. This proves that $\exists\, \mathcal{M} \subset \mathcal{L}$ which is invariant under the operator $\mathbf{T}_i$. Thus condition 2) of *Definition A.2.1* holds. Let us consider $\mathcal{G}_0(\mathbf{T}_i\omega) =$

$$\sum_{k=0}^{\infty}(\mathbf{T}_i\omega[k])^T M_0\mathbf{T}_i\omega[k] + \zeta_0 = \sum_{k=i}^{\infty}\omega[k-i]^T M_0\omega[k-i] + \zeta_0$$

$= \mathcal{G}_0(\omega)$. Similarly, it can be observed that $\mathcal{G}_j(\mathbf{T}_i\omega) = \mathcal{G}_j(\omega)\ \forall j = \{0,\dots,k\}$. Thus condition 3) of *Definition A.2.1* holds. Since we have shown that the functionals $-\mathcal{G}_0(\omega(\cdot)),\mathcal{G}_1(\omega(\cdot)),\dots,\mathcal{G}_k(\omega(\cdot))$ satisfy conditions $1), 2)$ and 3) of *Definition A.2.1*, they form an S-system. This concludes the proof. ■

Now we are ready to present the proof of *Theorem 4.4*.

Proof:

Step 1: [Problem reformulation] Using the hypograph formulation, (10) can be rewritten as

$$\sup_{\upsilon,a\in\ell_{2e}} \left\{\upsilon \;\middle|\; \begin{matrix} \frac{1}{N_2}\sum_{i=1}^{N_2}||y_{p,i}||^2_{\ell_2} \geq \upsilon, \\ ||y_{r,i}||^2_{\ell_2} \leq 1,\ x_i[\infty] = 0,\ \forall i \in \{1,\dots,N_2\}\end{matrix}\right\} \quad (20)$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3318194

From now, in the next two steps of the proof, we focus on the optimization problem (20) without the state constraints

$$\sup_{v, a \in \ell_{2e}} \left\{ v \;\middle|\; \begin{array}{l} \frac{1}{N_2}\sum_{i=1}^{N_2} ||y_{p,i}||_{\ell_2}^2 \geq v, \\ ||y_{r,i}||_{\ell_2}^2 \leq 1, \; \forall i \in \{1, \dots, N_2\} \end{array} \right\} \tag{21}$$

The reason to focus on (21) rather than (20) is that S-procedure becomes convenient to apply when there are no equality constraints. Thus, we drop the state constraints now and introduce them back at the end of *Step 3*. Equivalently, (21) be reformulated as

$$\inf_{v} \left\{ v \middle| \left\{ a \in \ell_{2e} \middle| \begin{array}{l} \frac{1}{N_2}\sum_{i=1}^{N_2} ||y_{p,i}||_{\ell_2}^2 - v > 0, \\ 1 - ||y_{r,i}||_{\ell_2}^2 \geq 0, \forall i \end{array} \right\} = \emptyset \right\} \tag{22}$$

Step 2: Let the system with the isolated uncertainty δ_i, with attack vector as input and performance and detection outputs as outputs be $\tilde{\Sigma}_{p,i} \triangleq (A_{cl,i}, B_{cl,i}, C_{p,i}, D_{p,i})$ and $\tilde{\Sigma}_{r,i} \triangleq (A_{cl}, B_{cl,i}, C_{r,i}, D_{r,i})$. Let us consider a linear time-invariant system of the form (18) with the attack vector a as input $\mu(\cdot)$ and the vector $\begin{bmatrix} y_{p,1} & y_{r,1}, \dots, y_{pN_2} & y_{r,N_2} \end{bmatrix}^T$ as output $\sigma(\cdot)$. This system will be stable due to *Assumption 2.1* and the system matrices of (18) would read

$$\left[\begin{array}{c|c} \Phi & \Lambda \\ \hline \Pi & \Upsilon \end{array}\right] = \left[\begin{array}{ccc|c} A_{cl,1} & 0 & 0 & B_{cl,1} \\ & \ddots & & \vdots \\ 0 & 0 & A_{cl,N_2} & B_{cl,N_2} \\ \hline C_{p,1} & 0 & 0 & D_{p,1} \\ C_{r,1} & 0 & 0 & D_{r,1} \\ & \ddots & & \vdots \\ 0 & 0 & C_{p,N_2} & D_{p,N_2} \\ 0 & 0 & C_{r,N_2} & D_{r,N_2} \end{array}\right],$$

For this system, let us define the set $\mathcal{L}$ as in (19) where $\omega = \begin{bmatrix} a^T & \sigma \end{bmatrix}^T \in \mathbb{R}^{n_a + N_2(n_p+n_r)}$. In view of (22), let us define

$$\mathcal{G}_0(\omega) \triangleq \frac{1}{N_2}\sum_{i=1}^{N_2} ||y_{p,i}||_{\ell_2}^2 - v = \sum_{k=0}^{\infty} \omega[k]^T M_0 \omega[k] + \zeta_0 \tag{23}$$

where $M_0 \in \mathbb{R}^{(n_a+N_2(n_p+n_r)) \times (n_a+N_2(n_p+n_r))}, \zeta_0 = -v$ and $M_0(i,j) = 1$, if $i = j, n_a + i$ is odd and 0 elsewhere. Similarly, let

$$\mathcal{G}_k(\omega) \triangleq -||y_{r,k}||_{\ell_2}^2 + 1 \;=\; \sum_{k=0}^{\infty} \omega[k]^T M_k \omega[k] + \zeta_k, \tag{24}$$

$\forall k = \{1, 2, \dots, N_2\}$, where $\zeta_k = -1$ and $M_k(i,j) = -1$, if $i = j, i = n_a + 2k - 1$ and 0 elsewhere. Here, M_k has same dimension as M_0 $\forall k = \{1, 2, \dots, N_2\}$. Therefore, we have shown that the constraints of (22) can be rewritten as functionals of the set $\mathcal{L}$. We can now see that the functionals $-\mathcal{G}_0(\cdot), \mathcal{G}_1(\cdot), \dots, \mathcal{G}_k(\cdot)$ along with *Lemma 2.1* satisfy the conditions under which *Theorem A.2.1* holds. Thus, by applying *Theorem A.2.1*, it follows that the functionals $-\mathcal{G}_0(\cdot), \mathcal{G}_1(\cdot), \dots, \mathcal{G}_k(\cdot)$ form an S-system. Let this be argument 1.

In the case the adversary chooses not to attack the system, i.e., $a = 0 \in \ell_{2e}$, it follows that $||y_{r,i}||_{\ell_2}^2 \approx 0 \; \forall \delta_i$. The residual energy $||y_{r,i}||_{\ell_2}^2$ is not strictly zero since there might be residual outputs due to the difference in initial condition between the system and the detector. The threshold ($\epsilon_r = 1$) is chosen in such a way that $||y_{r,i}||_{\ell_2}^2 \ll 1 \; \forall \delta_i$ when $a = 0$. Thus, it holds that $\exists \; \omega_0 = [a, \; \omega] = [0, \; 0_+]$ s.t. $-||y_{r,k}||_{\ell_2}^2 + 1 = \mathcal{G}_k(\omega_0) > 0 \; \forall k = \{1, \dots, N_2\}$. Here 0_+ represents a real number close to zero. Let this be argument 2. Finally, for any given $\omega \in \mathcal{L}$, let v^* be the corresponding optimal solution from (21). Then we know from (22) that the set of inequalities $\mathcal{G}_0(\omega) > 0, \mathcal{G}_i(\omega) \geq 0 \; i = \{1, \dots, k\}$ (where $\mathcal{G}_0(\omega)$ in (23) is constructed using the optimal v^*) is not solvable. Let this be argument 3. Using the above arguments (1-3), and [25, Theorem 4.3.1], we can conclude that, given the functionals in (23) and (24), there exists constants $\gamma_1 \geq 0, \gamma_1 \geq 0, \dots, \gamma_{N_2} \geq 0$ such that (25) holds

$$\mathcal{G}_0(\omega) + \sum_{i=1}^{N_2} \gamma_i \mathcal{G}_i(\omega) \leq 0, \; \forall \; \omega \in \mathcal{L}. \tag{25}$$

To conclude, in this step we have shown that (25) holds if the constraint of (22) holds. Additionally using [27, Theorem 1] (see also [25, Remark 4.3.1]), we observe that (25) implies that the constraints of (22) hold. As a result, we have that (22) and (25) are equivalent.

Step 3: We have shown that the constraint of (22) holds iff (25) is true. Then, we reformulate (22) as

$$\inf_{v, \gamma_1 \geq 0, \dots, \gamma_{N_2} \geq 0} \left\{ v \;\middle|\; \mathcal{G}_0(\omega) + \sum_{i=1}^{N_2} \gamma_i \mathcal{G}_i(\omega) \leq 0, \; \forall \; \omega(\cdot) \in \mathcal{L} \right\}.$$

Substituting the definition of $\mathcal{G}_0(\omega)$, we obtain

$$\inf_{\gamma \geq 0} \left\{ \inf_{v} \left\{ v \middle| \sum_{i=1}^{N_2} \left\{ \frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 \; + \gamma_i \mathcal{G}_i(\omega) \right\} \leq v, \forall \omega \right\} \right\} \tag{26}$$

where $\gamma = [\gamma_1, \dots, \gamma_{N_2}]^T$. The inner optimization problem of (26) resembles an epigraph formulation which can be rewritten as

$$\inf_{\gamma \geq 0} \left\{ \sup_{\omega} \; \left\{ \sum_{i=1}^{N_2} \left\{ \frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 \; + \gamma_i \mathcal{G}_i(\omega) \right\} \right\} \quad \right\}.$$

Substituting the definition of $\mathcal{G}_i(\omega), \; \forall i = \{1, \dots, N_2\}$, we obtain

$$\inf_{\gamma \geq 0} \underbrace{\sup_{\omega} \left\{ \sum_{i=1}^{N_2} \left\{ \frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 - \gamma_i ||y_{r,i}||_{\ell_2}^2 \right\} + \sum_{i=1}^{N_2} \gamma_i \right\}}_{\kappa} \tag{27}$$

Observe that κ is a maximization problem with a quadratic term in its objective. Thus, it holds that

$$\kappa = \begin{cases} \mathbf{1}^T \gamma, & \text{if } \; \sum_{i=1}^{N_2} \left\{ \frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 \; - \gamma_i ||y_{r,i}||_{\ell_2}^2 \right\} \leq 0 \\ +\infty, & \text{otherwise} \end{cases}.$$

Using the above result in (27), we obtain

$$\inf_{\gamma \geq 0} \; \left\{ \mathbf{1}^T \gamma \middle| \sum_{i=1}^{N_2} \left[\frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 - \gamma_i ||y_{r,i}||_{\ell_2}^2 \leq 0 \right], \forall a \in \ell_{2e} \right\}$$

where ω is replaced by a since it is the only control variable. Finally, we add the state constraints that were removed while formulating the optimization problem (21)

$$\inf_{\gamma \geq 0} \left\{ \mathbf{1}^T \gamma \middle| \begin{array}{l} \sum_{i=1}^{N_2} \left[\frac{1}{N_2} ||y_{p,i}||_{\ell_2}^2 - \gamma_i ||y_{r,i}||_{\ell_2}^2 \leq 0 \right], \forall a \\ x_i[\infty] = 0 \; \forall i \in \{1, \dots, N_2\} \end{array} \right\} \tag{28}$$

Thus in this step, we have shown using S-procedure that the optimization problem (20) and (28) are equivalent.

Step 4: Define $\bar{x}[\infty] = \begin{bmatrix} x_1[\infty]^T & \dots & x_{N_2}[\infty]^T \end{bmatrix}^T$, $\bar{x}[0] = \begin{bmatrix} x_1[0]^T & \dots & x_{N_2}[0]^T \end{bmatrix}^T$, $\bar{y}_p = \begin{bmatrix} y_{p1}^T & \dots & y_{pN_2}^T \end{bmatrix}^T$, and $\bar{y}_r = \begin{bmatrix} y_{r1}^T & \dots & y_{rN_2}^T \end{bmatrix}^T$. Using these definitions, the constraint of (28) can be rewritten as

$$-\frac{1}{N_2} ||\bar{y}_p||_{\ell_2}^2 + ||\sqrt{\Gamma(\gamma)} \bar{y}_r||_{\ell_2}^2 \geq 0, \forall a \in \ell_{2e}, \;\; \bar{x}[\infty] = 0 \tag{29}$$

where $\Gamma(\gamma)$ is defined in the theorem statement. Additionally due to *Assumption 2.3*, we have $\bar{x}[0] = 0$. Next let us define $y_1 = \sqrt{\Gamma(\gamma)}\bar{y}_r, y_2 = \frac{1}{\sqrt{N_2}}\bar{y}_p$ and the supply rate $s(\cdot) \triangleq ||y_1||_2^2 - ||y_2||_2^2$. Then, we have shown from (29) that [28, *Proposition 2, 2)*] holds. Equivalently, using [28, *Proposition 2, 3)*], we replace (29) by the constraint of (13) where $\bar{\Sigma}_p = (\bar{A}, \bar{B}, \bar{C}_p, \bar{D}_p)$ and $\bar{\Sigma}_r = (\bar{A}, \bar{B}, \bar{C}_r, \bar{D}_r)$ represent the system with the attack as input and $\bar{y}_p$ and $\bar{y}_r$ as system outputs respectively. Constructing these system matrices, as we did in *Step 2* of this proof concludes the proof. ■

A.3. Proof of *Lemma 4.5*

Proof: To recall, the optimization problem (13) was formulated using [28, *Proposition 2, 3)*] where $y_1 = \sqrt{\Gamma(\gamma)}\bar{y}_r$ and $y_2 = \frac{1}{\sqrt{N_2}}\bar{y}_p$. Here $\bar{y}_p$ and $\bar{y}_r$ represents the outputs of $\bar{\Sigma}_p$ and $\bar{\Sigma}_r$ respectively. Due to the equivalency between 3) and 4) of [28, *Proposition 2, 3)*], the FDI [28, *Proposition 2, 4)*] should hold $\forall\, |z| = 1$. Since we know that $y_1 = \sqrt{\Gamma(\gamma)}\bar{y}_r$ and $y_2 = \bar{y}_p$, we can deduce that $G_1(z)$ corresponds to $\sqrt{\Gamma(\gamma)}\bar{G}_r(z)$ and $G_2(z)$ to $\frac{1}{N_2}\bar{G}_r(z)$ in [28, *Proposition 2, 4)*], where $\bar{G}_r(z) = \bar{C}_r(z_1 I - \bar{A})^{-1}\bar{B} + \bar{D}_r$ and $\bar{G}_p(z) \triangleq \bar{C}_p(z_1 I - \bar{A})^{-1}\bar{B} + \bar{D}_p$. Thus, (13) can be rewritten as

$$\inf\left\{\mathbf{1}^T\gamma \middle| \bar{G}_r(\bar{z})^T\Gamma(\gamma)\bar{G}_r(z) - \bar{G}_p^T(\bar{z})\bar{G}_p(z) \succeq 0, \forall |z| = 1\right\} \quad (30)$$

Let us define the following sets such that $\mathbb{C}^{n_a} = \mathcal{Z}_{pr} \cup \mathcal{Z} \cup \mathcal{Z}_r \cup \mathcal{Z}_p$.

$$\begin{aligned}
\mathcal{Z}_{pr} &\triangleq \{x \in \mathbb{C}^{n_a} \mid \bar{G}_r(z)x = 0, \bar{G}_p(z)x = 0\},\\
\mathcal{Z} &\triangleq \{x \in \mathbb{C}^{n_a} \mid \bar{G}_r(z)x \neq 0, \bar{G}_p(z)x \neq 0\},\\
\mathcal{Z}_r &\triangleq \{x \in \mathbb{C}^{n_a} \mid \bar{G}_r(z)x = 0, \bar{G}_p(z)x \neq 0\},\\
\mathcal{Z}_p &\triangleq \{x \in \mathbb{C}^{n_a} \mid \bar{G}_r(z)x \neq 0, \bar{G}_p(z)x = 0\}.
\end{aligned}$$

Sufficiency: For any given z such that $|z| = 1$, if $x \in \mathcal{Z}_p$ or $x \in \mathcal{Z}_{pr}$, choosing $\Gamma(\gamma) = 0$ satisfies the constraint of (30). Similarly, if $x \in \mathcal{Z}$, let us pick $\Gamma(\gamma) = cI_{n_r}$ where c is a constant. Then, the value of (30) is bounded if there exists a bounded c that makes $\Xi \triangleq \sup_{|z|=1, x\in\mathcal{Z}} \frac{x^H\left\{\bar{G}_r^T(\bar{z})\bar{G}_r(z)\right\}x}{x^H\left\{\bar{G}_p^T(\bar{z})\Gamma(\gamma)\bar{G}_{p,i}(z)\right\}x}$ bounded. Ξ is bounded since the denominator cannot become zero (since $x \in \mathcal{Z}$ and $\Gamma(\gamma)$ is full rank), and we have assumed that the $\bar{G}_r(z)$ and $\bar{G}_p(z)$ are stable (*Assumption 2.1*). Next, we prove sufficiency when $x \in \mathcal{Z}_r$.

When condition 1) of the lemma statement holds, by definition of a zero $\forall |z| = 1, \nexists s \neq 0 \in \mathbb{C}^{n_a}$ such that $\bar{G}_r(z)s = 0$. Thus it follows that $\mathcal{Z}_r = \mathcal{Z}_{pr} = \emptyset$. When condition 2) of the lemma statement holds, by definition of a zero $\forall |z| = 1, \nexists s \neq 0$ such that $\bar{G}_r(z)s = 0$ and $\bar{G}_p(z)s \neq 0$. Thus it follows that $\mathcal{Z}_r = \emptyset$.

Necessity: Assume that there exists a bounded $\Gamma(\gamma)$ that solves (30). We also assume that there exists a complex number z_1 on the unit circle which is a zero of the system $\bar{\Sigma}_r$ (including multiplicity and input direction) but are not zeros of $\bar{\Sigma}_p$. By definition of a zero, it holds that $\exists s \neq 0$ such that $\bar{G}_r(z_1)s = 0$ and $\bar{G}_p(z_1)s \neq 0$. Under these assumptions, when $z = z_1$ and $x = s$, the constraint of (30) can be rewritten as $-s^H\bar{G}_p^T(\bar{z}_1)\bar{G}_{p,i}(z_1)s \geq 0$ which cannot hold since $\bar{G}_p(z_1)s \neq 0$. That is, the feasibility set of (30) is empty which contradicts our assumption. This concludes the proof. ■

References

[1] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.
[2] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid," *Electricity Information Sharing and Analysis Center*, vol. 388, 2016.
[3] K. E. Hemsley, E. Fisher, *et al.*, "History of industrial control system cyber incidents," tech. rep., Idaho National Lab, Idaho Falls, U.S., 2018.
[4] *Guide for conducting risk assessment.* N.I.S.T., 2012.
[5] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and commun. security*, pp. 355–366, 2011.
[6] Y. W. Law, T. Alpcan, and M. Palaniswami, "Security games for risk minimization in automatic generation control," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 223–232, 2014.
[7] J. Milošević, A. Teixeira, K. H. Johansson, and H. Sandberg, "Actuator security indices based on perfect undetectability: Computation, robustness, and sensor placement," *IEEE Trans. on Automat. Control*, 2020.
[8] C. Kwon and I. Hwang, "Hybrid robust controller design: Cyber attack attenuation for cyber-physical systems," in *52nd IEEE conference on decision and control*, pp. 188–193, IEEE, 2013.
[9] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, "Cyber risk analysis of combined data attacks against power system state estimation," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3044–3056, 2018.
[10] 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 Trans. on Automat. Contr.*, vol. 64, no. 12, pp. 4907–4919, 2019.
[11] S. Harshbarger, M. Hosseinzadehtaher, B. Natarajan, E. Vasserman, M. Shadmand, and G. Amariucai, "(A Little) ignorance is bliss: The effect of imperfect model information on stealthy attacks in power grids," in *2020 IEEE Kansas Power and Energy Conf.*, pp. 1–6, IEEE, 2020.
[12] C. Murguia, I. Shames, J. Ruths, and D. Nešić, "Security metrics and synthesis of secure control systems," *Automatica*, vol. 115, p. 108757, 2020.
[13] 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.
[14] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: the output-to-output ℓ_2-gain," in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 2582–2587, IEEE, 2015.
[15] S. C. Anand and A. M. Teixeira, "Joint controller and detector design against data injection attacks on actuators," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 7439–7445, 2020.
[16] S. X. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.
[17] S. M. Djouadi, A. M. Melin, E. M. Ferragut, J. A. Laska, J. Dong, and A. Drira, "Finite energy and bounded actuator attacks on cyber-physical systems," in *2015 European Control Conference (ECC)*, pp. 3659–3664, IEEE, 2015.
[18] S. D. Bopardikar, A. Speranzon, and J. P. Hespanha, "An h-infinity approach to stealth-resilient control design," in *2016 Resilience Week (RWS)*, pp. 56–61, IEEE, 2016.
[19] M. Ahmadi, X. Xiong, and A. D. Ames, "Risk-averse control via CVaR barrier functions: Application to bipedal robot locomotion," *IEEE Control Systems Letters*, vol. 6, pp. 878–883, 2021.
[20] K. W. Lye and J. M. Wing, "Game strategies in network security," *Intl. Journal of Information Security*, vol. 4, no. 1-2, pp. 71–86, 2005.
[21] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Mathematics of operations research*, vol. 23, no. 4, pp. 769–805, 1998.
[22] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized algorithms for analysis and control of uncertain systems: with applications*. Springer Science & Business Media, 2012.
[23] M. C. Campi, S. Garatti, and F. A. Ramponi, "A general scenario theory for nonconvex optimization and decision making," *IEEE Transactions on Automatic Control*, vol. 63, no. 12, pp. 4067–4078, 2018.
[24] A. Ruszczyński and A. Shapiro, "Optimization of convex risk functions," *Mathematics of operations research*, vol. 31, no. 3, pp. 433–452, 2006.
[25] I. R. Petersen, V. A. Ugrinovskii, and A. V. Savkin, *Robust Control Design Using H_∞ Methods*. Springer Science & Business Media, 2012.
[26] S. C. Anand, A. M. Teixeira, and A. Ahlén, "Risk assessment of stealthy attacks on uncertain control systems," *arXiv preprint arXiv:2106.07071v1*, 2021.
[27] V. Yakubovich, "Nonconvex optimization problem: The infinite-horizon linear-quadratic control problem with quadratic constraints," *Systems & Control Letters*, vol. 19, no. 1, pp. 13–22, 1992.
[28] A. M. Teixeira, "Optimal stealthy attacks on actuators for strictly proper systems," in *2019 IEEE 58th Conf. on Decision and Contr. (CDC)*, pp. 4385–4390, IEEE, 2019.