Optimal stealthy attacks on actuators for strictly proper systems

André M. H. Teixeira

***Abstract*— In this paper, we consider stealthy data injection attacks against control systems, and develop security sensitivity metrics to quantify their impact on the system. The final objective of this work is to use such metrics as objective functions in the design of optimal resilient controllers against stealthy attacks, akin to the classical design of optimal $\mathcal{H}_\infty$ robust controllers. As a first metric, the recently proposed ℓ_2 output to output gain is first examined, and fundamental limitations of this gain for systems with strictly proper dynamics are uncovered and characterized. To circumvent such limitations, a new security sensitivity metric is proposed, namely the truncated ℓ_2 gain. Necessary and sufficient conditions for this gain to be finite are derived, which we show can cope with strictly proper systems. Finally, we report preliminary investigations on the design of optimal resilient controllers, which are supported and illustrated through numerical examples.**

I. Introduction

The topic of cyber-secure control systems has been receiving increasing attention recently. An overview of existing cyber-threats and vulnerabilities in networked control systems is presented in [1], [2]. Rational adversary models are highlighted as one of the key items in security for control systems, thus making adversaries endowed with intelligence and intent, as opposed to faults. Therefore, these adversaries may exploit existing vulnerabilities and limitations in the traditional anomaly detection mechanisms and remain undetected. In fact, [3] uses tools from geometric control to study such fundamental limitations and characterizes a set of stealthy attack policies for networked systems modeled by differential-algebraic equations. Related stealthy attack policies were also considered in [2], [4], while the work by [5] characterizes the number of corrupted sensor channels that cannot be detected during a finite time-interval. A common thread within these approaches is that stealthy attacks are constrained to be entirely decoupled from the anomaly detector's output. Classes of attacks that are in theory detectable, but hard to detect in practice, have not received as much attention.

Another important direction is to analyze the potential damage of stealthy attacks. Recently, [6] investigated the detectability limitations and performance degradation of data injection attacks in stochastic control systems. The impact of stealthy data injection attacks on sensors is also investigated in [7], which characterized the set of states reachable by stealthy adversary. The work in [8] formulated the impact of data injection attacks in finite time-horizon as a generalized eigenvalue problem, whereas [9] considered an alternative formulation that allowed for the impact to be characterized as the solution to a convex optimization problem. A similar approach was considered in [10] for impulsive attacks.

While this set of results is useful to assess the impact of stealthy cyber-attacks on control systems, they cannot be used to directly design more resilient controllers, since the optimization problems have a complex non-convex dependence on the design parameters.

The impact and detectability of data injection attacks has also been jointly considered in the author's previous work [11], where the impact of stealthy attacks is characterized as the solution to a convex problem with linear matrix inequalities (LMIs). This characterization has a remarkable similarity with existing optimization-based techniques to design optimal $\mathcal{H}_\infty$ robust detectors and controllers [12], [13], which points to the possibility of using the metric in [11] to design resilient control systems.

As main contributions of this paper, we revisit the sensitivity metric developed in [11], and show that it has a fundamental limitation for analyzing systems with strictly proper dynamics from the attack signal to the anomaly detector's output. To circumvent such a limitation, an alternative metric is proposed, namely the truncated ℓ_2 output to output gain. Properties of this gain and its use in the design of optimal resilient controllers against stealthy attacks are investigated.

The outline of the paper is as follows. In Section II, we describe the problem formulation, present the closed-loop system under data injection attacks to the actuation signals, and describe the adversary model under consideration. A background on dissipative systems theory is provided in Section III, to support the discussions in the following sections. Section IV formulates the optimal attack policy, and analyzes fundamental limitations in such a formulation for systems with strictly proper parts. To address such limitations, a novel attack policy is proposed in Section V, which allows for strictly proper systems to be considered. Properties of this novel policy are investigated, and its use in the design of optimal controllers is discussed. In Section VI, the results are illustrated for a scalar system, for The paper concludes with final remarks and discussion on future work in Section VII.

A. Notation

Denote $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}$. and $\mathbb{Z}^+$ as the set of real, complex, integer, and non-negative integer numbers, respectively. The set of matrices with m rows, n columns, and entries in $\mathbb{R}$ ($\mathbb{C}$) is denoted as $\mathbb{R}^{m\times n}$ ($\mathbb{C}^{m\times n}$). A positive (semi-)definite square matrix $A \in \mathbb{C}^{n\times n}$ is denoted as $A \succ 0$ ($A \succeq 0$). Let $\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n$ be a real-valued discrete-time signal and

A. Teixeira is with the Division of Signals and Systems at the Uppsala University, Uppsala, Sweden. andre.teixeira@angstrom.uu.se

This work was supported in part by the Swedish Research Council (grant 2018-04396).

denote $x[k] \in \mathbb{R}^n$ as its value at time $k \in \mathbb{Z}^+$. Considering the time-horizon $[0, N] = \{k \in \mathbb{Z}^+ | \ 0 \leq k \leq N\}$ and the real-valued signals $\mathbf{x}$ and $\mathbf{y}$, denote the ℓ_2-norm of $\mathbf{x}$ over $[0, N]$ as $\|\mathbf{x}\|^2_{[0,N]}$. Let the space of square integrable signals be defined as as $\ell_2 \triangleq \{\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n | \ \|\mathbf{x}\|^2_{[0,\infty]} < \infty\}$ and define the extended signal space $\ell_{2e} \triangleq \{\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n | \ \|\mathbf{x}\|^2_{[0,N]} < \infty, \forall N \in \mathbb{Z}^+\}$.

II. Problem formulation and background

In this section, we present the model of the closed-loop system and characterize the adversary model.

A. Closed-loop control system

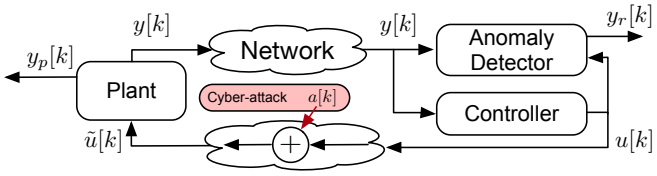

Fig. 1. Control system under a data injection attack on the control signal.

Consider the control system illustrated in Fig. 1, which is composed by a plant ($\mathcal{P}$), a controller ($\mathcal{C}$), and an anomaly detector ($\mathcal{D}$). The closed-loop system is under a cyber-attack on the control signal, where the adversary ($\mathcal{A}$) injects false-data on the control signal transmitted to the actuator. For the sake of simplifying the presentation, suppose that the controller is a static output feedback, and the anomaly detector is a state observer, described respectively by

$$\mathcal{P} : \begin{cases} x_p[k+1] = Ax_p[k] + B\tilde{u}[k] \\ y[k] = Cx_p[k] \\ y_p[k] = C_J x_p[k] + D_J \tilde{u}[k] \end{cases} \tag{1}$$

$$\mathcal{C} : \{u[k] = Ly[k] \tag{2}$$

$$\mathcal{D} : \begin{cases} \hat{x}_p[k+1] = A\hat{x}_p[k] + Bu[k] + Ky_r[k] \\ y_r[k] = y[k] - C\hat{x}_p[k], \end{cases} \tag{3}$$

where $\tilde{u}[k] \in \mathbb{R}^{n_u}$ is the control signal received by the actuator, $u[k] \in \mathbb{R}^{n_u}$ is the control signal computed by the controller, $y[k] \in \mathbb{R}^{n_m}$ is the measurement signal used by the controller and anomaly detector, $\hat{x}_p[k] \in \mathbb{R}^{n_x}$ is the state estimate, $y_p[k] \in \mathbb{R}^{n_p}$ is the performance output, and $y_r[k] \in \mathbb{R}^{n_r}$ is the detection output (residual) that is evaluated to ascertain the presence of anomalies. The controller gain is denoted by $L \in \mathbb{R}^{n_u \times n_m}$, while $K \in \mathbb{R}^{n_x \times n_m}$ is the observer gain. The closed-loop system is said to have a good performance if the energy of the performance output, $\|\mathbf{y_p}\|_{\ell_2}$, is small. Additionally, an anomaly is said to be detected if the energy of the detection output, $\|\mathbf{y_r}\|_{\ell_2}$, is larger than a given threshold.

Remark 1: In this formulation, the performance output $y_p[k]$ relates to typical quadratic costs considered in optimal control problems, where,

$$\|\mathbf{y_p}\|^2_{\ell_2} = \sum_{k=0}^{\infty} \begin{bmatrix} x[k] \\ \tilde{u}[k] \end{bmatrix}^\top \begin{bmatrix} C_J^\top \\ D_J^\top \end{bmatrix} \begin{bmatrix} C_J & D_J \end{bmatrix} \begin{bmatrix} x[k] \\ \tilde{u}[k] \end{bmatrix}$$

B. False-data injection attack scenario

Given the structure of the closed-loop system described above, we now present the attack scenario. In particular, we discuss the model knowledge and disruption and disclosure resources available to the adversary, together with the adversary's goals and constraints shaping the attack policy.

Disruption and disclosure resources: In the present scenario, the adversary can inject false-data in the actuator and channels, which is captured by having

$$\mathcal{A} : \{\tilde{u}[k] = u[k] + a[k]$$

where $a[k] \in \mathbb{R}^{n_u}$ is the data corruption inserted by the adversary. However, the adversary cannot eavesdrop on the sensor and actuator data. Hence, the corresponding attack policy does not use any online data of the system, corresponding to an open-loop policy, and is further assumed to be computed *a priori*.

Model knowledge: In the present scenario, the adversary also has access to the detailed model of the closed-loop system as described below, which is used to compute the attack policy. Defining $e[k] \triangleq x_p[k] - \hat{x}_p[k]$ and $x[k] \triangleq [x_p[k]^\top \ e[k]^\top]^\top$, the closed-loop system dynamics under attack, describing how the outputs $y_p[k]$ and $y_r[k]$ depend on the attack $a[k]$, are given by

$$\Sigma \triangleq \begin{cases} x[k+1] = A_{cl}x[k] + B_{cl}a[k] \\ y_p[k] = C_p x[k] + D_p a[k] \\ y_r[k] = C_r x[k] + D_r a[k] \end{cases}, \tag{4}$$

where the matrices are given by

$$\begin{aligned} A_{cl} &\triangleq \begin{bmatrix} A + BLC & 0 \\ 0 & A - KC \end{bmatrix}, & B_{cl} &\triangleq \begin{bmatrix} B \\ B \end{bmatrix} \\ C_p &\triangleq \begin{bmatrix} C_J + D_J LC & 0 \end{bmatrix}, & D_p &\triangleq D_J, \\ C_r &\triangleq \begin{bmatrix} 0 & C \end{bmatrix}, & D_r &\triangleq 0. \end{aligned} \tag{5}$$

Furthermore, we define the systems $\Sigma_p \triangleq (A_{cl}, B_{cl}, C_p, D_p)$ and $\Sigma_r \triangleq (A_{cl}, B_{cl}, C_r, D_r)$ for ease of presentation.

Attack goals and constraints: In this cyber-attack scenario, the adversary aims at designing the data corruption $a[k]$ as to maximize the impact on the performance output $y_p[k]$, while remaining undetected with respect to the detection output $y_r[k]$. The level of disruption, i.e., impact, is evaluated through the increase in the cost function $J_N(\mathbf{x_p}, \tilde{\mathbf{u}}) = \|\mathbf{y_p}\|^2_{[0,N]}$, the energy of the performance output. On the other hand, the adversary remains stealthy if no alarm is triggered, i.e., $\|\mathbf{y_r}\|^2_{[0,N]} \leq 1$. Thus the adversary is constrained to keep the energy of the detection output bounded. These elements lead to the following attack policy.

C. Strategic stealthy attack policy

Given the adversary model previously described, we let N go to infinity and consider optimal policies that maximize the energy of the performance output (i.e., the control cost) $J_\infty(\mathbf{x_p}, \tilde{\mathbf{u}}) = \|\mathbf{y_p}\|^2_{[0,\infty]}$ while ensuring that the energy of the detection output is bounded as $\|\mathbf{y_r}\|^2_{[0,\infty]} \leq 1$.

Therefore, the maximum ratio (gain) between $\|\mathbf{y_p}\|^2_{[0,\infty]}$ and $\|\mathbf{y_r}\|^2_{[0,\infty]}$ captures the maximum level of disruption

induced by a stealthy adversary, relative to the detection threshold. Such an attack policy will be further characterized and discussed in Section IV.

III. Dissipative systems theory

Consider the discrete-time system $\Sigma = (A, B, C, D)$ with state $x[k] \in \mathbb{R}^{n_x}$ and input $u[k] \in \mathbb{R}^{n_u}$. Define a real-valued function of the inputs and states of the system, called supply-rate, as $s : \mathbb{R}^{n_u} \times \mathbb{R}^{n_x} \to \mathbb{R}$, together with a non-negative function of the states $V : \mathbb{R}^{n_x} \to \mathbb{R}_+$, called storage function. In particular, we consider quadratic supply rates characterized by

$$s(u,x) = \begin{bmatrix} x \\ u \end{bmatrix}^\top \underbrace{\begin{bmatrix} Q_{xx} & Q_{xu} \\ Q_{ux} & Q_{uu} \end{bmatrix}}_{Q} \begin{bmatrix} x \\ u \end{bmatrix}, \tag{6}$$

where $Q = Q^\top \in \mathbb{R}^{n_x+n_u \times n_x+n_u}$, without any definiteness constraints being imposed on Q. Since Q is symmetric, and thus diagonalizable, note that Q can be decomposed as

$$Q = [C_r \; D_r]^\top [C_r \; D_r] - [C_p \; D_p]^\top [C_p \; D_p],$$

for appropriate matrices C_r, D_r, C_p, and D_p, and the supply rate can also be rewritten as $s(u,x) = \|y_r\|^2 - \|y_p\|^2$.

In the literature, the discrete-time system Σ is said to be dissipative with respect to the supply rate $s(u,x)$ if there exists a real-valued function $V(x)$ such that the inequality

$$\begin{aligned} V(x[k_1]) - V(x[k_0]) &\leq \sum_{k=k_0}^{k_1-1} s(u[k], x[k]) \\ &= \|\mathbf{y_r}\|^2_{[k_0,k_1]} - \|\mathbf{y_p}\|^2_{[k_0,k_1]} \end{aligned} \tag{7}$$

holds for all $k_0 \leq k_1$ and all trajectories of the system.

Remark 2: By writing the dissipation inequality in terms of a difference in between output energies, most of the definitions and results of dissipative systems for continuous-time systems can be straightforwardly mapped to discrete-time systems, and vice-versa, as it has been highlighted by different authors [14], [15]. Therefore, for brevity, the proofs in the present section are omitted.

Without loss of generality, for quadratic supply rates (6) and linear time-invariant systems, the storage functions can be taken as quadratic functions of the state of the form $V(x[k]) = x[k]^\top P x[k]$, with $P = P^\top$.

The next results, essential to the derivations presented in the next section, immediately follows from its continuous-term counter parts: [16, Theorems 8.4.5 and 8.4.9].

Proposition 1: Consider the LTI system $\Sigma = (A, B, C, D)$, which is assumed to be controllable, and the quadratic supply rate $s(u,x) = \|y_r\|^2 - \|y_p\|^2$. Define $G_r(z) = C_r(zI - A)^{-1}B + D_r$ and $G_p(z) = C_p(zI - A)^{-1}B + D_p$. The following statements are equivalent:

1) the system Σ is dissipative w.r.t. $s(u,x)$;
2) for all trajectories of the system such that $N > 0$ and $x[0] = 0$, we have $\sum_{k=0}^{N-1} s(x[k], u[k]) \geq 0$;
3) there exists a positive semi-definite matrix $P \succeq 0$ such that the following linear matrix inequality (LMI) holds:

$$\begin{bmatrix} A^\top P A - P & A^\top P B \\ B^\top P A & B^\top P B \end{bmatrix} - Q \preceq 0. \tag{8}$$

4) $G_r(\bar{z})^\top G_r(z) - G_p(\bar{z})^\top G_p(z) \succeq 0$ for all $z \in \mathbb{C}$ with $z \notin \sigma(A)$, $|z| \geq 1$.

Similar results can be established for the notion of cyclo-dissipative [15] (or internally dissipative [16]) systems.

Proposition 2: Consider the LTI system $\Sigma = (A, B, C, D)$, which is assumed to be controllable, and the quadratic supply rate $s(u,x) = \|y_r\|^2 - \|y_p\|^2$. Define $G_r(z) = C_r(zI - A)^{-1}B + D_r$ and $G_p(z) = C_p(zI - A)^{-1}B + D_p$. The following statements are equivalent:

1) the system Σ is cyclo-dissipative w.r.t. $s(u,x)$;
2) for all trajectories of the system such that $N > 0$ and $x[0] = x[N] = 0$, we have $\sum_{k=0}^{N-1} s(x[k], u[k]) \geq 0$;
3) there exists a symmetric matrix $P = P^\top$ such that the LMI (8) holds;
4) $G_r(\bar{z})^\top G_r(z) - G_p(\bar{z})^\top G_p(z) \succeq 0$ for all $z \in \mathbb{C}$ with $z \notin \sigma(A)$, $|z| = 1$.

IV. The ℓ_2 output to output gain

The shortcomings of the classical sensitivity metrics when applied to malicious attacks, as summarized in review of the the state-of-the-art, can be tackled by defining novel metrics that jointly trade-off the attack's impact and detectability. One instance of a metric tailored to consider the impact and detectability has been proposed in previous work [11], as summarized below.

Recall the example scenario described in Section II, where the closed-loop dynamics of the control system under attack are given by (4). Given an adversary that aims at maximizing the decrease in performance while remaining undetected, the corresponding attack policy can be formulated as the solution to the following non-convex optimization problem

$$\begin{aligned} \|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r} \triangleq \quad & \sup_{\mathbf{a} \in \ell_{2e}} && \|\mathbf{y_p}\|^2_{[0,\infty]} \\ & \text{subject to} && (4) \quad \forall k \geq 0, \quad x[0] = 0, \\ & && \|\mathbf{y_r}\|^2_{[0,\infty]} \leq 1, \end{aligned} \tag{9}$$

where $\|\Sigma\|^2_{\ell_2,\, y_p \leftarrow y_r}$ captures the maximum impact induced by a stealthy adversary.

Remark 3: Observe that non-vanishing undetectable attacks, which force $y_r[k] = 0$ for all k and have infinite energy, result in unbounded gain $\|\Sigma\|^2_{\ell_{2e},\, y_p}$.

The work in [11] first investigates necessary and sufficient conditions under which $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ admits a finite value, and later exploits methods from dissipative system theory to rewrite the non-convex problem (9) as a convex semidefinite programming problem, where $\|\Sigma\|^2_{\ell_{2e}, y_p \leftarrow y_r} = \gamma^\star$ is

obtained from the optimal solution of

$$\begin{array}{ll} \underset{P,\gamma}{\text{minimize}} & \gamma \\ \text{subject to} & P \succeq 0,\ \gamma > 0, \\ & R(\Sigma, P, \gamma) \preceq 0, \end{array} \tag{10}$$

where $R(\Sigma, P, \gamma)$ is given by

$$R(\Sigma, P, \gamma) \triangleq \begin{bmatrix} A_{cl}^\top P A_{cl} - P & A_{cl}^\top P B_{cl} \\ B_{cl}^\top P A_{cl} & B_{cl}^\top P B_{cl} \end{bmatrix} - \gamma \begin{bmatrix} C_r^\top \\ D_r^\top \end{bmatrix} \begin{bmatrix} C_r & D_r \end{bmatrix} + \begin{bmatrix} C_p^\top \\ D_p^\top \end{bmatrix} \begin{bmatrix} C_p & D_p \end{bmatrix}.$$

The work in [11] provides a first formulation and characterization of a novel sensitivity metric, $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$, that has a clear relation to cyber-security and quantifies the sensitivity of the system to malicious stealthy adversaries. However, as elaborated next, this metric suffers from a fundamental limitation when applied to systems with strictly proper components.

Lemma 1: Let $D_p \neq 0$ be of full column rank, and let $D_r = 0$. Then the ℓ_2 output to output gain gain is unbounded.

Proof: The proof follows directly from item 4) of Proposition 1, or alternatively from [11, Theorem 2], which states that the gain is finite if and only if the unstable zeros of Σ_r are contained in the zeros of Σ_p. For the case where $D_p \neq D_r = 0$, note that Σ_r has an unstable zero at infinity, for $z = \infty$. On the other hand, Σ_p does not have any zeros at infinity, which concludes the proof. ∎

A few remarks are in order. First, the latter fundamental limitation means that, for systems with $D_p \neq D_r = 0$, the metric $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ will always be unbounded, and thus it provides no other information about the susceptibility to stealthy attacks. Second, despite the existence of specific attacks that result in unbounded gains, we are still interested in assessing and mitigating the impact of other stealthy attacks. To overcome such a fundamental limitation and enable the analysis of more general stealthy attacks, we propose an alternative metric for assessing the impact of stealthy attacks.

V. The truncated ℓ_2 output to ouput gain

The limitation of $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ in handling systems with $D_r = 0$ and $D_p \neq 0$ may be overcome by looking into alternative formulations of the sensitivity metric that exclude the role of the zeros at infinity.

For instance, consider the scenario where the adversary wishes to maximize the impact while remaining undetected over a long period of time, and eventually stop the attack and leave without triggering an alarm. The impact of such an attack may be captured by the following sensitivity metric:

$$\begin{array}{lll} \|\Sigma\|^2_{\ell_{2e}[-\infty,\,0],\, y_p \leftarrow y_r} \triangleq & \underset{\mathbf{a} \in \ell_{2e}[-\infty,\,0]}{\sup} & \|\mathbf{y_p}\|^2_{[-\infty,\infty]} \\ & \text{subject to} & (4), \quad \forall k > -\infty, \\ & & x[-\infty] = 0, \\ & & \|\mathbf{y_r}\|^2_{[-\infty,+\infty]} \leq 1, \\ & & a[k] = 0,\ \forall k \geq 0. \end{array} \tag{11}$$

where the final constraint forces the attack to stop at $k = 0$. Note, however, that the impact and detection are considered even after the attack has ended, for $k \geq 0$.

Given that the attack stops at $k = 0$, the system for $k \geq 0$ becomes an autonomous system without external inputs, released from an initial condition $x[0]$ that depends on the past attack signal. Hence, $\|\mathbf{y_p}\|^2_{[0,\infty]}$ can be characterized as $\|\mathbf{y_p}\|^2_{[0,\infty]} = x[0]^\top W_p x[0]$, where W_p is the observability Gramian of the closed-loop system with respect to the performance output y_p. Similarly, we have $\|\mathbf{y_r}\|^2_{[0,\infty]} = x[0]^\top W_r x[0]$, where W_r is the observability Gramian of the closed-loop system with respect to the detection output y_r. Using these characterizations of $\|\mathbf{y_p}\|^2_{[0,\infty]}$ and $\|\mathbf{y_r}\|^2_{[0,\infty]}$, together with results from dissipative theory as in [11], the sensitivity metric (11) can then be re-written as

$$\begin{array}{lll} \|\Sigma\|^2_{\ell_{2e}[-\infty,\,0],\, y_p \leftarrow y_r} = & \underset{P = P^\top, \gamma > 0}{\text{minimize}} & \gamma \\ & \text{subject to} & P - W_p + \gamma W_r \succeq 0, \\ & & R(\Sigma, P, \gamma) \preceq 0. \end{array} \tag{12}$$

A. Necessary conditions for finite truncated gain

Necessary conditions under which the truncated gain is finite are now subject to analysis. First, note that the null spaces of $W_r \succeq 0$ and $W_p \succeq 0$ are related unobservable subspaces. As a consequence, the following result follows.

Lemma 2: The truncated gain is unbounded if there exists a subspace, associated with an unstable mode of the closed-loop system, that is reachable from the origin, unobservable for Σ_r, and observable for Σ_p.

Proof: The result follows straightforwardly from the time-domain characterization of the gain (11). First, recall that a state x results in $x^\top W_r x = 0$ if it is unobservable with respect to Σ_r. Second, if the same state is observable with respect to Σ_p, then it yields $x^\top W_p x \neq 0$. Hence, if there exists such a reachable subspace and it is associated with an unstable mode of A_{cl}, then an optimal attack policy would enforce $x[0] = \epsilon x$, for some small $\epsilon \neq 0$ such that the detectability constraint is met. As a result, the output $y_p[k]$ would be increasing after $k = 0$, while $y_r[k]$ would remain at constant, and thus result in an unbounded gain. ∎

The previous sufficient condition provides one instance when the gain is unbounded, which is related to the first LMI constraint in (12). However, it is not applicable to stable closed-loop systems. For the case of stable systems, we instead have the following result, which relates to the second constraint in (12).

Lemma 3: The truncated gain is unbounded if there exists a periodic trajectory from $x[-\infty] = 0$ to $x[0] = 0$ such that $\|y_r\|_{[-\infty,\,0]} = 0$ and $\|y_p\|_{[-\infty,\,0]} \neq 0$.

Proof: The result is a direct consequence of the time-domain characterization of the gain (11). Naturally, if the condition holds for a given attack signal, then such attack signal can be scaled arbitrarily and result in unbounded gain. ∎

Remark 4: As seen in Proposition 2, the concept of cyclo-dissipativity characterizes the non-negativity of the total

supply over all periodic trajectories, which in our case would mean that $\gamma\|y_r\|_{[-\infty,\,0]} - \|y_p\|_{[-\infty,\,0]} \geq 0$ must hold for periodic trajectories such that $x[-\infty] = x[0]$. In particular, a system is cyclo-dissipative if and only if $R(\Sigma, P, \gamma) \preceq 0$ holds for some symmetric matrix P, which is exactly the non-trivial constraint in (12). Hence, the absence of cyclo-dissipativity is indeed related to Lemma 3, and to the infeasibility of the second LMI constraint in (12).

In addition to these results, note that the gain is truncated gain is finite if both conditions are absent. In other words, a stable closed-loop system for which all periodic trajectories satisfy $\gamma\|y_r\|_{[-\infty,\,0]} - \|y_p\|_{[-\infty,\,0]} \geq 0$, for some $\gamma > 0$, is ensured to have a finite truncated gain. Thus the proposed gain successfully addresses the limitations of the original output to output gain for systems with $D_p \neq D_r = 0$.

Next we explore the problem of designing a controller that minimizes the truncated gain of the closed-loop system, for a fixed anomaly detector.

B. Design of resilient controllers

The security metrics formulated above allow to characterize the worst-case impact of a stealthy cyber-attack on the performance of the closed-loop system. Therefore, designing a controller that minimizes the security metrics would yield a reduced impact for the same level of detectability. The optimal controller L mitigating stealthy attacks can then be chosen as to minimize $\|\Sigma(L)\|^2_{\ell_{2e}[-\infty,\,0],\,y_p\leftarrow y_r}$, which may be formulated as the following optimization problem:

$$\begin{array}{ll} \underset{P=P^\top,\gamma,L}{\text{minimize}} & \gamma \\ \text{subject to} & P + \gamma W_r(L) - W_p(L) \succeq 0,\ \gamma > 0, \\ & R(\Sigma(L), \gamma) \preceq 0, \end{array} \tag{13}$$

where $W_r(L)$ and $W_p(L)$ are the observability Gramians of Σ_r and Σ_p, respectively, which also depend on the controller gain L through Lyapunov equations. Note that (13) is not a standard optimal $\mathcal{H}_\infty$ control problem, despite the possible resemblances. In particular, note that P, W_r, and W_p can all depend on the controller, and that P may not even be invertible, as required in standard $\mathcal{H}_\infty$ controller design techniques [13]. The solution to the above design problem therefore requires a closer investigation, which is left as future work. However, for the simple scalar case, one can evaluate the optimal controller by brute force. This exploration is reported in the next section through a numerical example.

VI. Numerical Example

In this section, we illustrate the results from earlier sections on the example system described in (4) with the following parameters: $A = 1.1$, $B = 0.1$, $C = 1$, $C_J = [1\ 0]^\top$, $D_J = [0\ 1]^\top$ $L = -10$ and $K = 0.5$.

A. ℓ_2 output to output gain

Constructing (4) with these parameters and solving (10), we observe that (10) is infeasible, and conclude that $\|\Sigma\|^2_{\ell_{2e},\,y_p\leftarrow y_r}$ is unbounded. The same result is obtained by using the method outlined in [8], where a finite-horizon version of (9), with a sufficiently large horizon $N \geq 1$, is solved using a generalized eigenvalue approach. A closer look reveals why the gain is unbounded: the system $(A_{cl}, B_{cl}, C_r, D_r)$ admits an unstable zero at infinity due to $D_r = 0$, while $(A_{cl}, B_{cl}, C_p, D_p)$ does not since $D_p \neq 0$. This corresponds precisely to result in Lemma 1.

B. Truncated ℓ_2 gain

Next, the novel sensitivity metric is illustrated on our example, together with the following finite-horizon version of the problem:

$$\begin{array}{lll} \|\Sigma\|^2_{\ell_{2e}[0,\,N],\,y_p\leftarrow y_r} \triangleq & \underset{\mathbf{a}\in\ell_{2e}[0,\,k_f]}{\sup} & \|\mathbf{y_p}\|^2_{[0,N]} \\ & \text{subject to} & (4),\ \forall k \geq 0,\ x[0] = 0, \\ & & \|\mathbf{y_r}\|^2_{[0,N]} \leq 1, \\ & & a[k] = 0,\ \forall k \geq k_f, \end{array} \tag{14}$$

where N and k_f are large and k_f is sufficiently smaller than N. In this example, we take $N = 200$ and $k_f = 150$. Solving the optimization problem (12) for our example, we obtain $\|\Sigma\|^2_{\ell_{2e}[-\infty,\,0],\,y_p\leftarrow y_r} = 935.14$. On the other hand, solving (14) through a generalized eigenvalue approach yields $\|\Sigma\|^2_{\ell_{2e}[0,\,N],\,y_p\leftarrow y_r} = 934.97$, based on which we conclude that the two formulations are in agreement and corroborate each other.

C. Optimal controller

Next, we show numerical results for simple cases that illustrate how the sensitivity metrics may be used as an objective function in the design of resilient controllers that minimize the impact of malicious stealthy cyber-attacks. Consider the system under attack as described in (4), and suppose that the adversary aims at maximizing the impact on the performance output y_p while remaining undetected with respect to the detection output y_r. In the following, we examine how the security metrics $\|\Sigma\|^2_{\ell_{2e},\,y_p\leftarrow y_r}$ and $\|\Sigma\|^2_{\ell_{2e}[-\infty,\,0],\,y_p\leftarrow y_r}$ may be used to design a controller that is resilient to such cyber-attacks for our illustrative example. Therefore, here we suppose that $K = 0.5$ is fixed, while the controller gain L is a decision variable.

Observe that A_{cl} and C_p (and thus Σ) depend linearly on the control gain L; to emphasize this, we shall use the notation $A_{cl}(L)$ and $C_p(L)$ (and $\Sigma(L)$). A natural approach to design resilient controllers is to choose L as to minimize one of the sensitivity metrics discussed earlier, as this results in a lower impact by stealthy attacks.

First we look into $\|\Sigma(L)\|^2_{\ell_{2e},\,y_p\leftarrow y_r}$ for $K = 0.5$. We observe through a line-search over the range of the stabilizing gains $L \in (-21,\ -1)$ that $\|\Sigma(L)\|^2_{\ell_{2e},\,y_p\leftarrow y_r}$ is always unbounded. Such an observation is expected, as the zeros of the system Σ are not modified by means of output feedback [17]. Second, we perform a line-search over the stabilizing control gains $L \in (-21,\ -1)$ for the one that minimizes $\|\Sigma\|^2_{\ell_{2e}[-\infty,\,0],\,y_p\leftarrow y_r}$. In fact, this may be achieved by iterating over $L \in (-21,\ -1)$ and, for each value of L, solve (12). For our illustrative example with $K = 0.5$, this method results in the choice $L = -1.4$

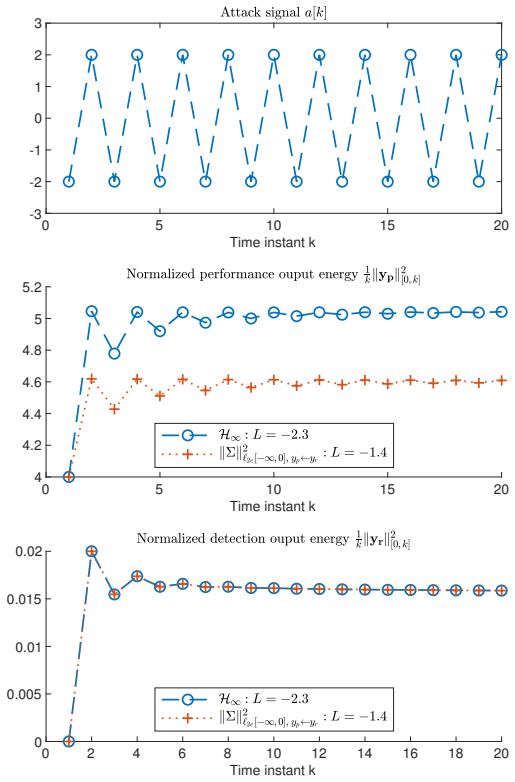

Fig. 2. Simulation example to highlight the difference between controllers.

and yields $\|\Sigma\|^2_{\ell_{2e}[-\infty,0],\, y_p \leftarrow y_r} = 294.54$. Compared to the initial controller with $L = -10$, the optimal controller $L = -1.4$ has a three times smaller worst-case impact by an undetectable attack. It is worth noting that an optimal $\mathcal{H}_\infty$ controller may be obtained in a similar way, by setting $C_r = 0$ and $D_r = 1$, performing a line-search over $L \in (-21,\ -1)$ and solving (10) with $P \succ 0$ for each L with the same system parameters, which results in the optimal $\mathcal{H}_\infty$ controller with $L = -2.3$. Fig. 2 illustrates the performance of the controllers $L = -1.4$ and $L = -2.3$ to the worst-case attack signal $a[k]$ (top plot) , where we observe that the respective systems have the same detection output energy (bottom plot), but the optimal $\mathcal{H}_\infty$ controller $L = -2.3$ has a higher performance output energy than the optimal resilient controller $L = -1.4$ (center plot), which indicates that the worst-case attack has a higher impact to the optimal $\mathcal{H}_\infty$ controller, for the same level of detectability.

The above numerical results further highlights the differences between using the proposed security metrics versus the classical metrics, such as the $\mathcal{H}_\infty$-norm, to design resilient controllers to mitigate possible stealthy cyber-attacks.

VII. Conclusions

In this paper, we analyzesd and developed novel sensitivity metrics that can jointly assess the impact and detectability of attacks. As a first metric, the recently proposed ℓ_2 output to output gain was first examined, and we show that, as a fundamental limitation, this gain is unbounded for systems with strictly proper dynamics with respect to the detection output, but with non-zero feed-through terms in the performance output. To circumvent such limitation, a new security sensitivity metric is proposed, namely the truncated ℓ_2 gain. Necessary and sufficient conditions for this gain to be finite are derived, which we show can cope with strictly proper systems. The final objective of this work is to use such metrics, which jointly consider impact and detectability, as objective functions in the design of optimal resilient controllers against stealthy attacks, akin to the classical design of optimal $\mathcal{H}_\infty$ robust controllers. Preliminary investigations on the design of optimal resilient controllers are reported, which are supported and illustrated through numerical examples.

Future work includes a deeper investigation of optimal controller design problem, as well as a more generic characterization of the fundamental limitation of the ℓ_2 gain for strictly proper systems.

References

[1] A. A. Cárdenas, S. Amin, and S. S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *First International Workshop on Cyber-Physical Systems*, June 2008.

[2] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, no. 1, pp. 135–148, 2015.

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

[4] R. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," in *18th IFAC World Congress*, 2011.

[5] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, June 2014.

[6] C. Bai, F. Pasqualetti, and V. Gupta, "Data-injection attacks in stochastic control systems: Detectability and performance tradeoffs," *Automatica*, vol. 82, pp. 251–260, Aug. 2017.

[7] Y. Mo and B. Sinopoli, "On the Performance Degradation of Cyber-Physical Systems Under Stealthy Integrity Attacks," *IEEE Trans. Automat. Contr.*, vol. 61, no. 9, pp. 2618–2624, Sep. 2016.

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

[9] D. Umsonst, H. Sandberg, and A. A. Cardenas, "Security analysis of control system anomaly detectors," in *2017 Am. Control Conf.* IEEE, May 2017, pp. 5500–5506. [Online]. Available: http://ieeexplore.ieee.org/document/7963810/

[10] I. Shames, F. Farokhi, and T. H. Summers, "Security analysis of cyber-physical systems using H2 norm," *IET Control Theory Appl.*, vol. 11, no. 11, pp. 1749–1755, 2017.

[11] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: the output-to-output l2-gain," in *54th IEEE Conference on Decision and Control (CDC)*, Osaka, Japan, Dec. 2015.

[12] J. Wang, G. Yang, and J. Liu, "An lmi approach to H- index and mixed H-/H_∞ fault detection observer design," *Automatica*, vol. 43, no. 9, pp. 1656–1665, 2007.

[13] C. Scherer and S. Weiland, "Linear matrix inequalities in control," in *The Control Systems Handbook: Control System Advanced Methods*, W. S. Levine, Ed. CRC Press, 2010.

[14] G. C. Goodwin and K. S. Sin, *Adaptive filtering prediction and control.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1984.

[15] P. Moylan, *Dissipative Systems and Stability*, 2014. [Online]. Available: ftp://ftp.pmoylan.org/papers/DissBook.pdf

[16] H. Trentelman and J. C. Willems, "The dissipation inequality and the algebraic Riccati equation," in *The Riccati Equation*, ser. Communications and Control Engineering Series, S. Bittanti, A. J. Laub, and J. C. Willems, Eds. Springer Berlin Heidelberg, 1991, pp. 197–242.

[17] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.