2015 IEEE 54th Annual Conference on Decision and Control (CDC)
December 15-18, 2015. Osaka, Japan

Strategic Stealthy Attacks: the output-to-output ℓ_2-gain

André Teixeira, Henrik Sandberg, Karl H. Johansson

***Abstract*— In this paper, we characterize and analyze the set of strategic stealthy false-data injection attacks on discrete-time linear systems. In particular, the threat scenarios tackled in the paper consider adversaries that aim at deteriorating the system's performance by maximizing the corresponding quadratic cost function, while remaining stealthy with respect to anomaly detectors. As opposed to other work in the literature, the effect of the adversary's actions on the anomaly detector's output is not constrained to be zero at all times. Moreover, scenarios where the adversary has uncertain model knowledge are also addressed. The set of strategic attack policies is formulated as a non-convex constrained optimization problem, leading to a sensitivity metric denoted as the output-to-output ℓ_2-gain. Using the framework of dissipative systems, the output-to-output gain is computed through an equivalent convex optimization problem. Additionally, we derive necessary and sufficient conditions for the output-to-output gain to be unbounded, with and without model uncertainties, which are tightly related to the invariant zeros of the system.**

I. Introduction

Resilience may be characterized as the ability to maintain acceptable levels of operation in the presence of abnormal conditions. It is an essential property in industrial control systems that are the backbone of several critical infrastructures, such as electric power systems, transport networks, and water and gas distributions networks.

The trend towards using pervasive and open-standard information technology (IT) systems, such as the Internet and SCADA communication protocols, results in control systems becoming increasingly vulnerable to malicious cyberthreats. In classical IT systems, potential vulnerabilities to malicious adversaries are tackled by ensuring the system's cybersecurity, which may be defined as the state of being protected against the unauthorized access, change, or destruction of electronic data and services. In fact, cybersecurity is a core requirement in information and communication technologies that are ubiquitous in modern societies. However, when applied to industrial control systems, the latter definition of cybersecurity does not capture essential features of control applications: system functionality and safe operation. More specifically, even insecure industrial systems must comply with safety requirements, while strict functionality or performance requirements on the system may render many cybersecurity mechanisms unusable. To address these operational requirements, resiliency must also be sought.

Not surprisingly, the complementarity of cybersecurity and resilience in industrial control systems is clearly visible in the basic assumptions and focus of these approaches. For instance, traditional cybersecurity does not consider the interdependencies between the physical components and the IT infrastructures. On the other hand, classical control-theoretic approaches to resilience (e.g., fault-tolerance) typically deal with independent disturbances and faults; thus they are not tailored to handle possibly colluding malicious cyberthreats. Theory and tools to analyze and build cyber-secure and resilient control systems are, therefore, lacking and in need to be developed.

The topic of cyber-secure and resilient control systems has been receiving increasing attention recently. An overview of existing cyberthreats and vulnerabilities in networked control systems is presented in [1]–[3]. Particularly, realistic and 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, [4] uses such fundamental limitations to characterize a set of stealthy attack policies for networked systems modeled by differential-algebraic equations. Related stealthy attack policies were also considered in [3], [5]. A common thread within these approaches is that stealthy attacks are constrained to be entirely decoupled from the anomaly detector's output. Such requirement may be overly stringent, since attacks yielding a sufficiently small output can also remain undetected. Moreover, relevant properties such as the impact or the strategic nature of such stealthy attack policies are not addressed in the literature.

As main contributions of this paper, we consider threat scenarios where malicious adversaries aim at maximizing the system's operational cost through false-data injection attacks, while remaining stealthy with respect to anomaly detectors. Specifically, the set of strategic stealthy false-data injection attacks are formulated as a non-convex optimization problem, which leads to a sensitivity metric denoted as the system's output-to-output ℓ_2-gain. Using the framework of dissipative systems, we propose computational methods to compute the output-to-output ℓ_2-gain. Additionally, we derive necessary and sufficient conditions for the output-to-output ℓ_2-gain to be unbounded, which are formulated as conditions on the system's invariant zeros. Furthermore, the existence of strategic attacks yielding unbounded gains under

This work has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 608224, the Swedish Research Council under Grants 2013-5523 and 2014-6282, the Swedish Foundation for Strategic Research, and the Knut and Alice Wallenberg Foundation.

A. Teixeira, H. Sandberg, K. H. Johansson are with the ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden. {andretei,hsan,kallej}@kth.se

model uncertainty is also analyzed.

The outline of the paper is as follows. In Section II, we describe the problem formulation and state the main questions that are tackled in this work. The dissipative systems theory for discrete-time systems is summarized in Section III, which will be used to derive one of the main results in Section IV. Conditions for the finiteness of the output-to-output gain with accurate models are derived in Section IV, which are illustrated through a simple numerical example. Section V tackles the robustness of the strategic attacks with respect to model uncertainty, and the paper concludes with final remarks in Section VI.

A. Notation

Denote $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}.$ and $\mathbb{Z}^+$ as the set of real, complex, integer, and positive 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 $A \succ 0$ ($A \succeq 0$). Given a pair of complex vectors $y, x \in \mathbb{C}^n$, denote their inner-product in $\mathbb{C}^n$ as $\langle y, x\rangle = y^H x$, where y^H is the Hermitian conjugate of y. The 2-norm of $x \in \mathbb{C}^n$ is defined as $\|x\| = \sqrt{\langle x, x\rangle}$. Let $\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n$ be a real-valued discrete-time signal and denote $x[k] \in \mathbb{R}^n$ as its value at time $k \in \mathbb{Z}^+$. Considering the time-horizon $[0, N] = \{k \in \mathbb{Z}^+ |\ 0 \le k \le N\}$ and the real-valued signals $\mathbf{x}$ and $\mathbf{y}$, denote the inner-product of $\mathbf{x}$ and $\mathbf{y}$ over $[0, N]$ as $\langle \mathbf{y}, \mathbf{x}\rangle_{[0,N]} = \sum_{k=0}^{N-1} \langle y[k], x[k]\rangle$. In particular, the ℓ_2-norm of $\mathbf{x}$ over $[0, N]$ is defined as $\|\mathbf{x}\|^2_{[0,N]} = \langle \mathbf{x}, \mathbf{x}\rangle_{[0,N]}$. Denote the space of square integrable signals as $\ell_2 = \{\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n |\ \|\mathbf{x}\|^2_{[0,\infty]} < \infty\}$ and define the extended signal space $\ell_{2e} = \{\mathbf{x} : \mathbb{Z}^+ \to \mathbb{R}^n |\ \|\mathbf{x}\|^2_{[0,N]} < \infty,\ \forall N \in \mathbb{Z}^+\}$.

II. Problem formulation

In this section, we present the control system structure and describe the main problem at hand. Consider the modeling framework described in [3], where the control system is composed by a physical plant, a feedback controller, and an anomaly detector. The physical plant, feedback controller, and anomaly detector are modeled in a discrete-time state-space form as, respectively,

$$\mathcal{P} : \begin{cases} x_p[k+1] = A_p x_p[k] + B_p \tilde{y}_c[k] \\ y_m[k] = C_m x_p[k] \\ y_p[k] = C_J x_p[k] + D_J \tilde{y}_c[k] \end{cases} \tag{1}$$

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

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

where $x_p[k] \in \mathbb{R}^{n_p}$, $z[k] \in \mathbb{R}^{n_z}$, and $s[k] \in \mathbb{R}^{n_s}$ are the state variables, $\tilde{y}_c[k] \in \mathbb{R}^{n_c}$ is the vector of control actions applied to the process, $y_m[k] \in \mathbb{R}^{n_m}$ is the measurement vector obtained from the sensors, $y_p[k] \in \mathbb{R}^{n_p}$ is a virtual output vector used to compute the closed-loop performance, and $y_r[k] \in \mathbb{R}^{n_r}$ the residue vector. The sensor measurements and actuator data are transmitted through a communication network, which at the plant side correspond to $y_m[k]$ and $\tilde{y}_c[k]$, respectively. At the controller side, we denote the sensor and actuator data by $\tilde{y}_m[k] \in \mathbb{R}^{n_y}$ and $y_c[k] \in \mathbb{R}^{n_c}$.

The controller is designed to optimize the closed-loop system's performance. Given the system trajectories within the time-interval $[0,\ N]$, the system's performance is evaluated according to the cost function $J_N(\mathbf{x_p}, \tilde{\mathbf{y}}_\mathbf{c}) = \|C_J \mathbf{x_p} + D_J \tilde{\mathbf{y}}_\mathbf{c}\|^2_{[0,N]} = \|\mathbf{y_p}\|^2_{[0,N]}$, where $\mathbf{y_p}$ is the virtual performance signal defined by $y_p[k] = C_J x_p[k] + D_J \tilde{y}_c[k]$.

The anomaly detector monitors the system to detect possible anomalies, i.e., deviations from the nominal behavior. The anomaly detector is collocated with the controller and it evaluates the behavior of the plant based only on the closed-loop models, $\tilde{y}_m[k]$ and $y_c[k]$. In particular, given the time-interval $[0,\ N]$ and the residue signal $\mathbf{y_r}$, an alarm is triggered to indicate the presence of anomalies if the residue meets $\|\mathbf{y_r}\|^2_{[0,\ N]} \ge \varepsilon^2$, where $\varepsilon \ge 0$ is chosen according to a suitable trade-off between detection and false-alarm rates. Without loss of generality, let $\varepsilon = 1$ in the remainder of this paper.

A. 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 measurement channels, which is captured by having

$$\begin{bmatrix} \tilde{y}_c[k] \\ \tilde{y}_m[k] \end{bmatrix} = \begin{bmatrix} y_c[k] \\ y_m[k] \end{bmatrix} + \begin{bmatrix} B_a \\ D_a \end{bmatrix} u[k],$$

where $u[k] \in \mathbb{R}^{n_u}$ is the attack vector. 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: Stacking the states of the plant, controller, and anomaly detector as $x[k] = [x_p[k]^\top \ \ z[k]^\top \ \ s[k]^\top]^\top$, the closed-loop dynamics under attack can be written as

$$\Sigma \triangleq \begin{cases} x[k+1] = Ax[k] + Bu[k] \\ \underbrace{\begin{bmatrix} y_p[k] \\ y_r[k] \end{bmatrix}}_{y[k]} = \underbrace{\begin{bmatrix} C_p \\ C_r \end{bmatrix}}_{C} x[k] + \underbrace{\begin{bmatrix} D_p \\ D_r \end{bmatrix}}_{D} u[k], \end{cases} \tag{4}$$

where $x[k] \in \mathbb{R}^{n_x}$, $y[k] \in \mathbb{R}^{n_y}$, and the individual matrices

are given by

$$
\begin{aligned}
A &= \begin{bmatrix} A_p + B_p D_c C_m & B_p C_c & 0 \\ B_c C_m & A_c & 0 \\ (B_e D_c + K_e) C_m & B_e C_c & A_e \end{bmatrix}, \\
B &= \begin{bmatrix} B_p B_a + B_p D_c D_a \\ B_c D_a \\ (B_e D_c + K_e) D_a \end{bmatrix}, \\
C_p &= \begin{bmatrix} C_J + D_J D_c C_m & D_J C_c & 0 \end{bmatrix}, \\
D_p &= D_J (D_c D_a + B_a), \\
C_r &= \begin{bmatrix} (D_e D_c + E_e) C_m & D_e C_c & C_e \end{bmatrix}, \\
D_r &= (D_e D_c + E_e) D_a.
\end{aligned} \tag{5}
$$

In the present scenario, the adversary also has access to the detailed model of the closed-loop system, $\Sigma = (A, B, C, D)$, which is used to compute the attack policy. Later in the paper, we also consider a more moderate scenario, where the adversary's knowledge contains some uncertainties.

Attack goals and constraints: The adversary aims at disrupting the system's behavior while remaining stealthy with respect to the anomaly detector. The level of disruption is evaluated through the increase in the cost function $J_N(\mathbf{x_p}, \tilde{\mathbf{y}}_\mathbf{c}) = \|\mathbf{y_p}\|^2_{[0,N]}$, while the adversary remains stealthy if no alarm is triggered, i.e., $\|\mathbf{y_r}\|^2_{[0,N]} \leq 1$. In particular, we let N go to infinity and consider adversaries that desire to maximize the cost $J_\infty(\mathbf{x_p}, \tilde{\mathbf{y}}_\mathbf{c}) = \|\mathbf{y_p}\|^2_{[0,\infty]}$ while ensuring that the residue output is bounded as $\|\mathbf{y_r}\|^2_{[0,\infty]} \leq 1$. Such an adversary model leads to the attack policy characterized next.

B. Strategic stealthy attack policy

Given the adversary model previously described, the corresponding attack policy can be formulated as the following non-convex optimization problem

$$
\begin{aligned}
\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r} \triangleq \quad & \sup_{\mathbf{u} \in \ell_{2e}} && \|\mathbf{y_p}\|^2_{[0,\infty]} \\
& \text{subject to} && (4), \quad \forall k \geq 0, \quad x[0] = 0, \\
& && \|\mathbf{y_r}\|^2_{[0,\infty]} \leq 1,
\end{aligned} \tag{6}
$$

where $\|\Sigma\|^2_{\ell_2,\, y_p \leftarrow y_r}$ captures the maximum level of disruption induced by a stealthy adversary. The resemblance of the optimization problem (6) with the classical input-to-output ℓ_2-gain of Σ is evident: simply replace $\mathbf{y_r}$ with $\mathbf{u}$ in (6) to obtain the input-to-output ℓ_2-gain. Such similarities compel us to denote $\|\Sigma\|^2_{\ell_2,\, y_p \leftarrow y_r}$ as the output-to-output ℓ_2-gain of Σ. In fact, the latter term alludes to an interesting re-interpretation of the output-to-output gain of Σ as the classical input-to-output gain of another system related to Σ. This interpretation carries valuable insights on important properties of $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ that are briefly discussed next, such as conditions under which the gain can be unbounded.

Consider the infinite-dimensional vector $\mathbf{y}_r \triangleq [y_r^\top[0] \;\; y_r^\top[1] \;\; \dots]^\top$ and define the infinite-dimensional Toeplitz operator $\mathcal{T}_r$ such that $\mathbf{y}_r = \mathcal{T}_r \mathbf{u}$ for $x[0] = 0$. With a slight abuse of notation, define $\mathcal{T}_r^\dagger$ as a left-inverse mapping of $\mathcal{T}_r$, such that $\mathbf{u} = \mathcal{T}_r^\dagger \mathbf{y}_r$. Using this transformation, the output-to-output ℓ_2-gain defined in (6) can be rewritten as the eigenvalue problem

$$
\begin{aligned}
\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r} = \quad & \sup_{\mathbf{y}_r \in \ell_{2e}} && \mathbf{y}_r^\top \mathcal{T}_r^{\dagger\top} \mathcal{T}_p^\top \mathcal{T}_p \mathcal{T}_r^\dagger \mathbf{y}_r \\
& \text{subject to} && \mathbf{y}_r^\top \mathbf{y}_r \leq 1,
\end{aligned} \tag{7}
$$

which corresponds to the classical ℓ_2-gain of the system $\mathcal{T}_p \mathcal{T}_r^\dagger$ with input y_r and output y_p. Although this interpretation may seem odd at first, it leads to an interesting motivation from the framework of behavioral systems [6]. In behavioral system, instead of inputs and outputs, $u[k]$ and $y[k]$ are seen as signals generated from the set of allowed trajectories of a system. In this sense, the output-to-output ℓ_2-gain simply represents the maximum amplification from one signal, $y_r[k]$, to another, $y_p[k]$. Moreover, we see that the system $\mathcal{T}_p \mathcal{T}_r^\dagger$ must be well-defined and asymptotically stable for $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ to be finite. In fact, we shall conclude that (A, B, C_r, D_r) having no unstable zeros outside the unit disk, thus admitting a left-inverse, is sufficient for the output-to-output gain to be finite. With additional derivations, necessary and sufficient conditions for the finiteness of $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ are derived. Specifically, in the subsequent sections we tackle the following questions:

1) How can $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ be computed?
2) What are the necessary and sufficient conditions for $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ to be bounded?
3) How sensitive are strategic stealthy attacks to model uncertainties?

In the remainder of this paper, we analyze relevant properties of the output-to-output ℓ_2-gain and address the previous questions. In particular, the first and second questions are tackled in Section IV, while the third question is addressed in Section V. First, we revisit the basic concepts of dissipative systems with quadratic supply-rates [7], [8], which are used to derive the main results in the paper.

III. Dissipative systems theory

Consider the discrete-time system Σ, as detailed in (4). 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

$$
s(u, x) = \begin{bmatrix} x \\ u \end{bmatrix}^\top \underbrace{\begin{bmatrix} Q_{xx} & Q_{xu} \\ Q_{ux} & Q_{uu} \end{bmatrix}}_{Q} \begin{bmatrix} x \\ u \end{bmatrix}, \tag{8}
$$

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

$$
s(u, x) = \|y_r\|^2 - \|y_p\|^2. \tag{9}
$$

In the literature, the discrete-time system (4) 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

$$\begin{aligned} V(x[k_1]) - V(x[k_0]) &\leq \sum_{k=k_0}^{k_1-1} s(u[k],x[k]) \\ &= \|\mathbf{y_r}\|^2_{[k_0,k_1]} - \|\mathbf{y_p}\|^2_{[k_0,k_1]} \end{aligned} \tag{10}$$

holds for all $k_0 \leq k_1$ and all trajectories satisfying (4).

Remark 1: 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 [9], [10]. Therefore, for brevity, the proofs in the present section are omitted.

For dissipative systems, there exist two universal storage functions of special interest, namely the available storage function $V^a_{[k_0,\,k_1]}(x)$ and the required supply function $V^r_{[k_0,\,k_1]}(x)$. The available storage, corresponding to the maximum supply extracted from the system (4) initialized at a fixed initial condition x_0, is defined as

$$\begin{aligned} V^a_{[k_0,\,k_1]}(x) \triangleq \quad & \sup_{\mathbf{u}\in\ell_{2e}} && -\sum_{k=k_0}^{k_1-1} s(u[k],y[k]) \\ & \text{subject to:} && (4),\ \forall k_0 \leq k \leq k_1, \\ & && x[k_0] = x,\ x[k_1] = 0. \end{aligned} \tag{11}$$

Subtracting k_0 to the time variables, defining $N = k_1 - k_0$, and using (9) yields

$$\begin{aligned} V^a_{[0,\,N]}(x) \triangleq \quad & \sup_{\mathbf{u}\in\ell_{2e}} && -(\|\mathbf{y_r}\|^2_{[0,N]} - \|\mathbf{y_p}\|^2_{[0,N]}) \\ & \text{subject to:} && (4),\ \forall 0 \leq k \leq N, \\ & && x[0] = x,\ x[N] = 0. \end{aligned} \tag{12}$$

The required supply function is defined as

$$\begin{aligned} V^r_{[k_0,\,k_1]}(x) \triangleq \quad & \inf_{\mathbf{u}\in\ell_{2e}} && \sum_{k=k_0}^{k_1-1} s(u[k],y[k]) \\ & \text{subject to:} && (4),\ \forall k_0 \leq k \leq k_1, \\ & && x[k_0] = 0,\ x[k_1] = x, \end{aligned} \tag{13}$$

which corresponds to the minimum supply needed to drive the system from the origin to a given final state. For later convenience, we rewrite the required supply function when time goes from 0 to N. Changing the time variables by subtracting k_0, defining $N = k_1 - k_0$, and using (9) yields

$$\begin{aligned} V^r_{[0,\,N]}(x) \triangleq \quad & \inf_{\mathbf{u}\in\ell_{2e}} && \|\mathbf{y_r}\|^2_{[0,N]} - \|\mathbf{y_p}\|^2_{[0,N]} \\ & \text{subject to:} && (4),\ \forall 0 \leq k \leq N, \\ & && x[0] = 0,\ x[N] = x. \end{aligned} \tag{14}$$

A third storage function is of particular interest in this work, namely the free end-point available storage function [7], which is defined as

$$\begin{aligned} V^a_{f,[0,\,N]}(x) \triangleq \quad & \sup_{\mathbf{u}\in\ell_{2e}} && -(\|\mathbf{y_r}\|^2_{[0,N]} - \|\mathbf{y_p}\|^2_{[0,N]}) \\ & \text{subject to:} && (4),\ \forall 0 \leq k \leq N, \\ & && x[0] = x, \end{aligned} \tag{15}$$

where the end-point $x[N]$ is free. Given these three storage functions, the following relations are known.

Lemma 1: The following relation holds for any storage function $V_{[0,\,N]}(x)$:

$$V^a_{[0,\,N]}(x) \leq V_{[0,\,N]}(x) \leq V^r_{[0,\,N]}(x).$$

In addition, for any non-negative storage function $V_{[0,\,N]}(x)$, the following also holds:

$$0 \leq V^a_{f,[0,\,N]}(x) \leq V_{[0,\,N]}(x) \leq V^r_{[0,\,N]}(x).$$

Proof: The proofs in the present paper have been omitted due to space limitations and may be found in [11]. ■

Without loss of generality, for quadratic supply rates (8) 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 result, essential to the derivations presented in the next section, immediately follows from Lemma 1 and the definition of dissipative system.

Proposition 1: Consider the LTI system Σ described in (4), which is assumed to be controllable and dissipative with respect to the quadratic supply rate $s(u,x)$. The following statements are equivalent:

1) the free end-point available storage function $V^a_{f,[0,\,N]}$ exists, i.e., $+\infty > V^a_{f,[0,\,N]} \geq 0$;
2) there exists a positive semi-definite matrix $P \succeq 0$ such that the following linear matrix inequality (LMI) holds:

$$\begin{bmatrix} A^\top PA - P & A^\top PB \\ B^\top PA & B^\top PB \end{bmatrix} - Q \preceq 0. \tag{16}$$

□

IV. Strategic stealthy attacks: a dissipative systems approach

With the framework of dissipative discrete-time systems at hand, let us revisit the strategic attack policy (6). In particular, the following result characterizes the optimal value of (6) in terms of the existence of a positive semi-definite storage function.

Theorem 1: Consider the LTI system Σ described in (4) and the strategic attack policy (6). The optimal value of the strategic attack policy is given by $\|\Sigma\|^2_{\ell_{2e},y_p\leftarrow y_r} = \gamma^\star$, where $\gamma^\star$ is the solution to the convex optimization problem

$$\begin{aligned} \min_{P,\gamma} \quad & \gamma \\ \text{s.t.} \quad & P \succeq 0,\ \gamma > 0, \\ & R(P,\Sigma) - \gamma[C_r D_r]^\top [C_r D_r] + [C_p D_p]^\top [C_p D_p] \preceq 0, \end{aligned} \tag{17}$$

with

$$R(P,\Sigma) = \begin{bmatrix} A^\top PA - P & A^\top PB \\ B^\top PA & B^\top PB \end{bmatrix}. \tag{18}$$

□

A. Unbounded output-to-output gain

We now derive necessary and sufficient conditions for $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ to have an unbounded value. To do so, we first characterize a set of necessary and sufficient conditions on the output signals $\mathbf{y_p}$ and $\mathbf{y_r}$, which are later restated as conditions on Σ.

Consider the inequality constraint of (6), i.e., $\|\mathbf{y}_r\|^2_{[0,\infty]} \leq 1$, and note that this inequality implies that $\mathbf{y}_r$ belongs to the ℓ_2 signal space, for which a necessary condition is that $y_r[\infty] = 0$. On the other hand, for $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ to be unbounded, a necessary condition is for $\mathbf{y_p}$ not to belong to ℓ_2, that is, $\|\mathbf{y_p}\|_{\ell_2} = +\infty$. Together, these two conditions are necessary and sufficient for $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ to be unbounded, as formalized in the following statement.

Lemma 2: Consider the system Σ. The value of $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ is unbounded if and only if, for any scalar $\epsilon > 0$, there exists an input signal $\mathbf{u} \in \ell_{2e}$ and integer $N \geq 0$ such that the following inequalities hold:

1) $\|\mathbf{y_r}\|_{[N,\infty]} \leq \epsilon$;
2) $\|\mathbf{y_p}\|_{[N,\infty]} \geq \epsilon^{-1}$.

□

As a preliminary step, the following lemma characterizes a set of necessary conditions on $\mathbf{x}$ and $\mathbf{y_r}$, which are useful to establish results in terms of Σ.

Lemma 3: Consider the system Σ. The value of $\|\Sigma\|^2_{\ell_{2e},\, y_p \leftarrow y_r}$ can only be unbounded if, for any scalar $\epsilon > 0$, there exists an input signal $\mathbf{u} \in \ell_{2e}$ and integer $N \geq 0$ such that the following inequalities hold:

1) $\|\mathbf{y_r}\|_{[N,\infty]} \leq \epsilon$;
2) $\|\mathbf{x}\|_{[N,\infty]} \geq \epsilon^{-1}$.

□

In fact, the existence of an input $\mathbf{u} \in \ell_{2e}$ satisfying the conditions in Lemma 3 with $\epsilon = 0$ and a sufficiently large N can be precisely characterized in terms of (A, B, C_r, D_r). The following definition is required for such characterization.

Definition 1: Consider a discrete-time linear time-invariant system with the state-space realization (A, B, C, D) and the equation

$$\begin{bmatrix} \lambda I - A & -B \\ C & D \end{bmatrix} \begin{bmatrix} x_\lambda \\ g_\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{19}$$

with $\lambda_z \in \mathbb{C}$ and $x_\lambda \neq 0$. For a given solution to the previous equation $(\lambda, g_\lambda, x_\lambda)$, denote λ as the invariant zero, g_λ as the input-zero direction, and x_λ as the state-zero direction. Furthermore, the tuple $(\lambda, g_\lambda, x_\lambda)$ is denoted as a zero-tuple of the system (A, B, C, D). Additionally, we denote a tuple $(\lambda, g_\lambda, x_\lambda)$ satisfying (20) with $|\lambda| \geq 1$ as an unstable zero-tuple, or simply unstable zero.

Lemma 4: Consider the system Σ with $x[0] = 0$. There exists an input signal $\mathbf{u} \in \ell_{2e}$ satisfying the conditions of Lemma 3 with $\epsilon \geq 0$ if and only if there exist a non-zero reachable state $x[N] = x_\lambda$, a complex vector $g_\lambda \in \mathbb{C}^{n_u}$, and a complex number $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$ satisfying

$$\begin{bmatrix} \lambda I - A & -B \\ C_r & D_r \end{bmatrix} \begin{bmatrix} x_\lambda \\ g_\lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{20}$$

□

Note that the input characterized in Lemma 4, if it exists, also satisfies Lemma 3 for all $\epsilon > 0$. Furthermore, it prescribes a class of attack signals that behave as two-stage attacks: first drive the system from the $x[0] = 0$ to $x[N] = x_\lambda$, then use $u[k] = \lambda^{k-N} g_\lambda$ for $k \geq N$. Using lemmas 2, 3, and 4, we now derive the following necessary and sufficient conditions for $\|\Sigma\|^2_{\ell_{2e}, y_p \leftarrow y_r}$ to be finite.

Theorem 2: Consider the LTI system Σ described in (4) and the strategic attack policy (6). The optimal value of the strategic attack policy is finite if and only if either of the following conditions hold:

1) the system (A, B, C_r, D_r) has no unstable zeros associated with a reachable x_λ;
2) the unstable zeros of the system (A, B, C_r, D_r) associated with a reachable x_λ are also zeros of (A, B, C_p, D_p).

□

Theorem 2 indicates that zero-dynamics attacks are indeed strategic in the sense of the attack policy (6), since they lead to unbounded gains. A standing open-question is whether such conclusion extends to the case where the adversary's knowledge is not accurate, which is addressed in the Section V. Before dealing with model uncertainty, we illustrate the earlier results through a simple numerical example.

B. Numerical Example

Consider the system $\Sigma = (A, B, C, D)$ with

$$\begin{aligned} &A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, C_r = \begin{bmatrix} 1 & 1 \end{bmatrix}, D_r = 1, \\ &C_p = \begin{bmatrix} 2+\delta & 2 \end{bmatrix}, D_p = 0, \end{aligned} \tag{21}$$

for some $\delta \in \mathbb{R}$. We are interested in characterizing the output-to-output ℓ_2-gain of the system for different values of δ. First, note that the system (A, B, C_r, D_r) has two zero-tuples, namely $(\lambda_s, x_{\lambda_s}, g_{\lambda_s}) = (0, [1 \;\; 1], -2)$ and $(\lambda, x_\lambda, g_\lambda) = (2, [1 \;\; -1]^\top, 0)$. The second tuple is the only unstable zero, so it is the only candidate to yield an unbounded gain. Since the system is controllable, all conditions of Lemma 4 are met. In fact, the candidate input signal $\mathbf{u}$ is given by $u[0] = 1$, $u[1] = -1$, and $u[k] = \lambda^{k-2} g = 0$ for all $k \geq 2$, for which the system reaches $x[2] = x_\lambda$ and then follows the zero-dynamics for $k \geq 2$. The existence of such input signal implies that condition 1) of Theorem 2 does not hold. Therefore, the finiteness of $\|\Sigma\|^2_{\ell_{2e}, y_p \leftarrow y_r}$ depends only on the second condition of Theorem 2, which we next analyze for different values of δ.

From condition 2) of Theorem 2, the gain $\|\Sigma\|^2_{\ell_{2e}, y_p \leftarrow y_r}$ is finite if and only if $(\lambda, x_\lambda, g_\lambda)$ is also a zero of $\Sigma_p = (A, B, C_p, D_p)$. From Definition 1 and having $(\lambda, x_\lambda, g_\lambda)$ as a zero of Σ_r, we observe that the latter condition is equivalent to have $[C_p D_p][x_\lambda^\top \; g_\lambda^\top]^\top = 0$, which only holds for $\delta = 0$. Thus, we conclude that $\|\Sigma\|^2_{\ell_{2e}, y_p \leftarrow y_r}$ is finite if and only if $\delta = 0$. In fact one can verify that, for $\delta = 0$, the optimization problem (17) has the optimal solution $(P^\star, \gamma^\star)$, with $P^\star = 4[1 \;\; 1]^\top [1 \;\; 1]$ and $\gamma^\star = 4$. In the present

example, one can easily verify that the candidate input signal $\mathbf{u}$ actually yields such gain, thus being the optimal input even when $\|\Sigma\|^2_{\ell_{2e},y_p\leftarrow y_r}$ is finite.

V. Strategic stealthy attacks with uncertainty

The example in the previous section indicates that the strategic attacks are robust to certain model uncertainties, in the sense that they yield unbounded gains even in the presence of uncertainties. In this section, we formally tackle this matter and characterize the robustness of strategic attacks w.r.t uncertainties. In particular, we relax the adversary's model knowledge and suppose that the adversary only has access to a set of uncertain linear systems, Ω_Σ, where a given system $\tilde{\Sigma} = (\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}) \in \Omega_\Sigma$ is of the form

$$\tilde{\Sigma} \triangleq \begin{cases} x[k+1] = \tilde{A}x[k] + \tilde{B}u[k] \\ \underbrace{\begin{bmatrix} y_p[k] \\ y_r[k] \end{bmatrix}}_{y[k]} = \underbrace{\begin{bmatrix} \tilde{C}_p \\ \tilde{C}_r \end{bmatrix}}_{\tilde{C}} x[k] + \underbrace{\begin{bmatrix} \tilde{D}_p \\ \tilde{D}_r \end{bmatrix}}_{\tilde{D}} u[k], \end{cases} \tag{22}$$

where

$$\begin{bmatrix} \tilde{A} & \tilde{B} \\ \tilde{C} & \tilde{D} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} + \begin{bmatrix} \Delta A & \Delta B \\ \Delta C & \Delta D \end{bmatrix}, \tag{23}$$

with (A, B, C, D) being the nominal system and $(\Delta A, \Delta B, \Delta C, \Delta D) \in \Omega_\Delta$ denoting the model uncertainty. Additionally, let Ω_Δ define the uncertainty set of interest, which is assumed to be closed and bounded and to include the zero uncertainty $(\Delta A, \Delta B, \Delta C, \Delta D) = (0, 0, 0, 0)$.

Given the uncertain set of linear systems Ω_Σ, we define the robust strategic attack policy as

$$\|\Omega_\Sigma\|^2_{\ell_2,\, y_p\leftarrow y_r} \triangleq \inf_{\tilde{\Sigma}\in\Omega_\Sigma} \|\tilde{\Sigma}\|^2_{\ell_{2e},y_p\leftarrow y_r} \tag{24}$$

Theorem 3: Consider the class of linear systems Ω_Σ described in (22) and the robust strategic attack policy (24). The optimal value of the robust strategic attack policy is unbounded if and only the system (A, B, C_r, D_r) has a non-empty set of unstable zeros, denoted as $\mathcal{Z}_r$, and there exists a tuple $(\lambda, g_\lambda, x_\lambda) \in \mathcal{Z}_r$ satisfying the following conditions:

1) $(\lambda, g_\lambda, x_\lambda) \in \mathcal{Z}_r$ is also a zero of all $\tilde{\Sigma}_r \in \Omega_{\Sigma_r}$;
2) x_λ is reachable from the origin for all $\tilde{\Sigma} \in \Omega_\Sigma$;
3) $(\lambda, g_\lambda, x_\lambda) \in \mathcal{Z}_r$ is not a zero of any $\tilde{\Sigma}_p \in \Omega_{\Sigma_p}$.

□

The latter result can be used to derive algebraic conditions on the model uncertainties such that the robust strategic attack policy yields finite values.

Lemma 5: The robust strategic attack policy yields an finite value if and only if there exists a given uncertainty $(\Delta A, \Delta B, \Delta C, \Delta D) \in \Omega_\Delta$ such that the equality

$$\begin{bmatrix} C_p + \Delta C_p & D_p + \Delta D_p \end{bmatrix} \begin{bmatrix} x_\lambda \\ g_\lambda \end{bmatrix} = 0 \tag{25}$$

holds for all $[x_\lambda^\top \; g_\lambda^\top]^\top$ such that:

1) x_λ is reachable from the origin;
2) the equality

$$\begin{bmatrix} \Delta A & \Delta B \\ \Delta C_r & \Delta D_r \end{bmatrix} \begin{bmatrix} x_\lambda \\ g_\lambda \end{bmatrix} = 0 \tag{26}$$

holds for all $(\Delta A, \Delta B, \Delta C_r, \Delta D_r) \in \Omega_{\Delta_r}$;
3) $(\lambda, g_\lambda, x_\lambda) \in \mathcal{Z}_r$, for some $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1$.

□

Since algebraic necessary and sufficient conditions for $\|\Omega_\Sigma\|^2_{\ell_2,\, y_p\leftarrow y_r}$ to be finite are given in Lemma 5, one may conclude that stealthy attacks yielding unbounded disruption levels are, in fact, quite sensitive to unstructured model uncertainty. As a particular example, define the model uncertainty set as $\Omega_\Delta(\delta) = \{(\Delta A, \Delta B, \Delta C, \Delta D) |\; \Delta B = \Delta C = \Delta D = 0, \Delta A = \delta I\}$. It is straightforward to verify that condition (26) in Lemma 5 does not hold for an arbitrarily small $\delta > 0$ and $x_\lambda \neq 0$, from which we conclude that $\|\Omega_\Sigma\|^2_{\ell_2,\, y_p\leftarrow y_r}$ is finite.

VI. Conclusions

In this work, considered threat scenarios where malicious adversaries inject false-data in order to maximize the system's operational cost, while remaining stealthy with respect to anomaly detectors. Specifically, the set of strategic stealthy false-data injection attacks were characterized as a non-convex optimization problem that lead to a system sensitivity metric denoted as the output-to-output gain. Computational methods to obtain this gain were proposed using the theory of dissipative systems. Necessary and sufficient conditions for the output-to-output gain to be unbounded were derived, which were formulated as properties of the system's invariant zeros. In addition, the effect of model uncertainty was also analyzed. Interesting future directions include the design of controllers and anomaly detectors yielding reduced output-to-output gains.

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. A. Cárdenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-physical Systems Security*. U.S. DHS, July 2009.

[3] 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.

[4] 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.

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

[6] J. C. Willems and J. W. Polderman, *Introduction to Mathematical Systems Theory: A Behavioral Approach*, ser. Texts in Applied Mathematics. Springer-Verlag New York, 1998.

[7] 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.

[8] N. Kottenstette, M. J. McCourt, M. Xia, V. Gupta, and P. J. Antsaklis, "On relationships among passivity, positive realness, and dissipativity in linear systems," *Automatica*, vol. 50, no. 4, pp. 1003–1016, 2014.

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

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

[11] A. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: the output-to-output ℓ_2-gain," *ArXiv e-prints*, 2015.