Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Joint controller and detector design against data injection attacks on actuators ⋆

Sribalaji C. Anand * **André M. H. Teixeira** *

* *Department of Electrical Engineering, Uppsala University, PO Box 534, SE-75121, Uppsala, Sweden (e-mail: {sribalaji.anand, andre.teixeira}@angstrom.uu.se)*

Abstract: This paper addresses the issue of data injection attacks on actuators in control systems. Considering attacks that aim at maximizing impact while remaining undetected, the paper revisits the recently proposed output-to-output gain, which is compared to classical sensitivity metrics such as H_∞ and H_-. In its original formulation, the output-to-output gain is unbounded for strictly proper systems. This limitation is further investigated and addressed by modifying the performance output of the system and ensuring that the system from attack signal to performance output is also strictly proper. With this system description, and by using the theory of dissipative systems, a Bi-linear Matrix Inequality (BMI) is formulated for system design. Using this BMI, a design algorithm is proposed based on the heuristic of alternating minimization. Through numerical simulations of the proposed algorithm, it is found that the output-to-output gain presents advantages over the other metrics: the effect of the attack is reduced in the performance output and increased in the detection output in a relatively large spectrum of frequencies.

Keywords: System security, Quadratic performance indices, Fault detection, H_∞ control, Optimization.

1. INTRODUCTION

The trend towards increased usage of open-standard communication protocols among industrial control systems has made these systems vulnerable to online cyber-attacks such as Stuxnet (Langner, 2011). The issue of cyber-attacks has been addressed in detail for classical Information Technology (IT) systems (Bishop, 2002). In IT systems, cyber-security deals with properties such as confidentiality, integrity, and availability. Although these properties are essential for control systems, other key features such as stability and safe operation are not addressed. Hence the results from classical IT security cannot be directly extended to control systems.

Security of control systems has been studied in detail from different contexts such as (a) Modelling of various possible attacks, (b) Detection of attacks, (c) Quantifying the impact of attacks and (d) Prevention and treatment of attacks (Chong et al., 2019).

Possible attack scenarios such as eavesdropping attack, denial-of-service attack, replay attack, bias injection attack, zero dynamics attack are described in Cárdenas et al. (2011). A common thread in these scenarios is that adversaries are considered to be rational, with given objectives, resources, and constraints. Detection techniques of attacks was studied for data injection attacks (Teixeira et al., 2012), replay attacks (Mo et al., 2015) and routing attacks (Ferrari and Teixeira, 2017). The context of the attack undetectability was studied in Pasqualetti et al. (2015). The impact caused by the aforementioned attacks on control system was studied in Milošević et al. (2018b) and Urbina et al. (2016). Detectability and impact of attacks are key aspects in the security of control systems since they characterize the robustness/vulnerability of the control system against attacks. Given that the control system is under attack, attack treatment/mitigation through secure state estimation both in continuous-time (CT) and discrete-time (DT) has been studied in Fawzi et al. (2014).

Nonetheless, there are still significant gaps in the existing literature. First, most papers have focused on mitigating sensor attacks, while stealthy attacks on actuators have not been as much investigated (Ye and Luo, 2019). Second, most of the work combining detection and impact has focused on system analysis (Milošević et al., 2018a), and these approaches are not amenable to design controllers and detectors for increased security. Third, the joint design of controllers and detectors has received little attention, partly due to the decoupled nature of the sensitivity metrics used in the related literature (Tan and Patton, 2015), (Ding et al., 2002).

This paper addresses the above mentioned research gaps by investigating the joint design of controllers and detectors against stealthy attacks on actuators. The contribution of this article is as follows: Firstly, we look into the general DT control system representation and investigate the shortcoming faced by certain sensitivity metrics when applied to strictly proper systems (Teixeira, 2019). Secondly, we look into a different approach to address this limitation in the following way: when control systems are sampled from CT to DT, certain classes of systems end

⋆ This work was supported by the Swedish Research Council under the grant 2018-04396.

up having algebraic loops which are not entirely causal, as supported by Blachuta (1999). Hence, we approach the sensitivity metrics with an altered system description (which respects causality) which in principle also addresses the aforementioned shortcomings of the sensitivity metric. Finally, we leverage the recent sensitivity metric, the output-to-output gain (OOG), and the proposed system description to cast the joint detector and controller design as an optimization problem with BMI constraints, which is tackled through an alternating minimization approach.

We conclude this section by providing the notations that are used in the paper. A formal problem background is provided is Section 2. Thereafter, a design problem based on sensitivity metrics is formulated in Section 3. The shortcoming faced by this design problem is discussed in Section 4. An altered system description is proposed to address this shortcoming. A design algorithm to the design problem based on this altered system description is presented. Section 5 provides numerical examples comparing the algorithm with the classical sensitivity metrics. Concluding remarks are provided in Section 6.

Notation:

Throughout this article, $\mathbb{R}, \mathbb{C}, \mathbb{Z}$ and $\mathbb{Z}^+$ represent the set of real numbers, complex numbers, integers and non-negative integers respectively. A positive (semi-)definite matrix A is denoted by $A \succ 0, (A \succeq 0)$. The maximum and minimum singular values of a matrix A is denoted by $\bar{\sigma}(A)$ and $\underline{\sigma}(A)$ respectively. The set of eigenvalues of a matrix A is represented by $\lambda(A)$. 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 integrable signals be defined as $\ell_2 \triangleq \{x : \mathbb{Z}^+ \to \mathbb{R}^n |\ ||x||^2_{\ell_2} \triangleq ||x||^2_{\ell_2,[0,\infty]} < \infty\}$ and the extended signal space be defined as $\ell_{2e} \triangleq \{x : \mathbb{Z}^+ \to \mathbb{R}^n |\ ||x||^2_{\ell_2,[0,N]} < \infty, \forall N \in \mathbb{Z}^+\}$. $0_{m\times n}(1_{m\times n})$ represents a matrix of size $m \times n$ where all the entries are zero (one).

2. PROBLEM BACKGROUND

In this section, we describe the control system structure and the goal of the stealthy adversary. Consider the general description of a closed-loop DT linear time-invariant system with a plant ($\mathcal{P}$), output feedback controller ($\mathcal{C}$) and anomaly detector ($\mathcal{D}$). For the sake of simplicity, we assume a static output feedback controller (2). The closed-loop system is represented 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 $x_p[k] \in \mathbb{R}^{n_x}$ is the state of the plant, $\tilde{u}[k] \in \mathbb{R}^{n_u}$ is the control signal applied to the actuator, $u[k] \in \mathbb{R}^{n_u}$ is the control signal generated by the controller, $y[k] \in \mathbb{R}^{n_m}$ is the measurement output produced by the plant, $y_p[k] \in \mathbb{R}^{n_p}$ is the virtual performance output, $\hat{x}[k] \in \mathbb{R}^{n_x}$ is the state estimate produced by the observer based detector, $y_r[k] \in \mathbb{R}^{n_m}$ is the residue generated by the detector, L and K are the controller and detector gains respectively. In general, the system is considered to have a good performance when the energy of the performance output ($||y_p||^2_{\ell_2}$) is small and an anomaly is considered to be detected when the energy of the residue ($||y_r||^2_{\ell_2}$) is greater than a predefined threshold (say ϵ_r). Without loss of generality, we assume $\epsilon_r = 1$ in the rest of this paper.

2.1 Data injection attack scenario

In the closed-loop system described above, we consider that an adversary is trying to inject false data into the actuator of the plant. Given this setup, we now discuss the resources the adversary has access to.

Disruption resources: The adversary can inject data on all the control channels. This is represented by:

$$\tilde{u}[k] \triangleq u[k] + a[k],$$

where $a[k] \in \mathbb{R}^{n_u}$ is the data injected by the adversary.

Model knowledge: The adversary has full system knowledge. This system knowledge is used by the adversary to calculate the optimal data injection attacks. Defining $e[k] \triangleq x_p[k] - \hat{x}_p[k]$ and $x[k] \triangleq [x_p[k]^T\ e[k]^T]^T$, the closed-loop system under attack with the performance output and detection output as system outputs becomes:

$$\mathcal{P}_{cl} : \begin{cases} x[k+1] &= A_{cl}x[k] + B_{cl}a[k] \\ y_p[k] &= C_p x[k] + D_J a[k] \\ y_r[k] &= C_r x[k] + D_r a[k], \end{cases} \tag{4}$$

where

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

Attack goals and constraints: The adversary aims at deteriorating the system performance while remaining undetected. Hence, the adversary injects attack signals to maximize the energy of the performance output while keeping the energy of the detection output lower than ϵ_r. This objective can be translated into an attack policy, which is formulated as the following optimization problem:

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

where $||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r}$ represents the OOG.

2.2 Dissipative systems theory

The OOG resulting from the optimization problem (5) can be used for capturing the disruption induced by an attack signal. This optimization problem (5) is non-convex and can be reformulated to its convex dual counterpart as:

$$\begin{aligned} ||\Sigma||^2_{\ell_{2e}, y_p \leftarrow y_r} \triangleq \min_{\gamma \geq 0} \ & \gamma \\ \text{s.t.} \quad & ||y_p||^2_{\ell_2} \preceq \gamma ||y_r||^2_{\ell_2} \ \ \forall\ a \in \ell_{2e}. \end{aligned} \tag{6}$$

The strong duality between (5) and the above optimization problem was shown to be zero in Teixeira et al. (2015). One drawback of the above optimization problem is that, it operates in signal space which is infinite dimensional. It is advantageous, if the optimization problem is in the form of finite dimensional matrix inequalities. This transformation can be done by dissipative systems theory. For brevity, a detailed presentation of dissipative system theory is omitted (interested reader can refer to Teixeira (2019) and references therein). But one of the important propositions is presented below to keep the presentation self-contained.

Definition 1. A system $\Sigma \triangleq (A, B, C, D)$ is said to be dissipative with respect to the supply function $s(\cdot)$ if there exists a real valued function $V(\cdot)$ such that the following equation holds:

$$V(x[k_1]) - V(x[k_0]) \leq \sum_{k=k_0}^{k_1-1} s(x[k], u[k]) \quad \forall k_1 \geq k_0$$

Remark 2. Without any loss of generality, for this setting, the storage function can be taken of the form $V(x[k]) = x[k]^T P x[k]$, with $P = P^T$.

Proposition 3. Consider a DT system Σ which is assumed to be controllable and observable. Let $y_i[k] = C_i x[k] + D_i u[k]$, $i = \{1, 2\}, s(x, u) = ||y_1[k]||_2^2 - ||y_2[k]||_2^2$. Then the following statements are equivalent:

(1) The system Σ is dissipative w.r.t $s(x, u)$
(2) For all trajectories of the system with $N > 0$ and $x[0] = 0$, it holds that $\sum_{k=0}^{N-1} s(x[k], u[k]) \geq 0$
(3) There exists a $P \succeq 0$ such that

$$\begin{bmatrix} A^T P A - P & A^T P B \\ B^T P A & B^T P B \end{bmatrix} - Q \preceq 0, \tag{7}$$

where $Q = [C_1 \ D_1]^T [C_1 \ D_1] - [C_2 \ D_2]^T [C_2 \ D_2]$.

Remark 4. Define $G_1(z) = C_1(zI - A)^{-1}B + D_1$ and $G_2(z) = C_2(zI - A)^{-1}B + D_2$. Under these definitions, the necessary condition for dissipativity (cyclo-dissipativity) is given by the following inequality:

$$G_1(\bar{z})^T G_1(z) - G_2(\bar{z})^T G_2(z) \succeq 0, \quad \forall z \in \mathbb{C} \text{ with } z \notin \lambda(A), |z| = 1$$

2.3 Output-to-output gain

Let $s(\cdot) \triangleq \gamma ||y_r||_{\ell_2}^2 - ||y_p||_{\ell_2}^2$. Under this definition, using (7), the squared OOG ($||\Sigma||_{\ell_{2e}, y_p \leftarrow y_r}^2$) which is the optimal value of (6), can be obtained by:

$$\min_{P \succeq 0, \gamma \geq 0} \quad \gamma$$
$$R(P) + \begin{bmatrix} C_p^T \\ D_p^T \end{bmatrix} [C_p \ D_p] - \gamma \begin{bmatrix} C_r^T \\ D_r^T \end{bmatrix} [C_r \ D_r] \preceq 0, \tag{8}$$

where $R(P) \triangleq \begin{bmatrix} A_{cl}^T P A_{cl} - P & A_{cl}^T P B_{cl} \\ B_{cl}^T P A_{cl} & B_{cl}^T P B_{cl} \end{bmatrix}$. Here, the optimal value of the optimization problem denotes the squared-*OOG*, which represents the disruption induced by the attack vector (optimizer of (5)) on the system.

3. DESIGN BASED ON SENSITIVITY METRICS

Sensitivity metrics can be defined as metrics quantifying the influence of an anomaly (attack in this case) signal on the system. One of the sensitivity metrics, the *OOG*, which captures the worst-case effect of the anomaly in the detection and performance output, was introduced in the previous section. Currently, there are a few other approaches in the literature that deals with defining impact metrics for the system under attack in the following ways: the H_∞ norm captures the worst-case (highest) effect of an anomaly in the performance output of the system, the H_- index captures the worst-case (lowest) effect of an anomaly in the detection output of the system.

These sensitivity metrics can be used as an objective function to design the controller and/or detector of the system by minimizing the impact on performance and maximizing the detectability of the anomaly in the system outputs. This section is aimed at describing the sensitivity metric and the corresponding design problems.

3.1 Output-to-Output gain (OOG) based design approach

The *OOG* can be used as an objective function to find a controller (L) and detector (K) such that the effect of the attack signal is increased in the detection output and reduced in the performance output simultaneously. The *OOG* optimal controller and detector is obtained by solving the optimization problem (9), where $||\Sigma||_{\ell_{2e}, y_p \leftarrow y_r} = \gamma^*$ is the worst-case ratio of the effect of the attack signal on the performance output to the detection output.

$$\min_{P, \gamma, L, K} \gamma^*$$
$$\text{s.t. } P \succeq 0, \ \gamma > 0$$
$$R(P, L, K) + \begin{bmatrix} C_p^T \\ D_p^T \end{bmatrix} [C_p \ D_p] - \gamma \begin{bmatrix} C_r^T \\ D_r^T \end{bmatrix} [C_r \ D_r] \preceq 0 \tag{9}$$

where $R(P, L, K) \triangleq \begin{bmatrix} A_{cl}^T P A_{cl} - P & A_{cl}^T P B_{cl} \\ B_{cl}^T P A_{cl} & B_{cl}^T P B_{cl} \end{bmatrix}$ and the matrices A_{cl} and C_p are functions of L and K.

Remark 5. The main advantage of using such a metric is that: from *Remark 4*, it is evident that the detector ($G_r(z)$) and the controller ($G_p(z)$) concentrate on the same frequency (z). This means that the attack detection and robustification occur at the same frequency, as opposed to the classical H_∞ and H_- methods. Hence the *OOG* design objective focuses on improving the detectability only when the impact of the attack signal on the performance output is sufficiently high at the same frequency. In other words, in frequency regions where the performance output is low, the *OOG* design problem (9) does not focus to improve the detectability. These observations are later illustrated through a numerical example in Section 5.

3.2 Classical H_∞ and H_- metrics based design approach

The objective of the H_∞ design approach is to find a feedback controller L, so as to reduce the effect of the disturbance/attack signal in the performance output. The classical H_∞ design optimization problem is:

$$||\Sigma||_{H_\infty}^2 \triangleq \min_{L} \quad \max_{a \in \ell_{2e}} \quad ||y_p||_{\ell_2}^2$$
$$\text{s.t.} \quad ||a||_{\ell_2}^2 = 1, \quad x[0] = 0 \tag{10}$$

where the optimal value of the H_∞ norm is $||\Sigma||_{H_\infty}$. The BMI approach for the optimization problem (10) exists in the literature (Hilhorst et al., 2014) and is given as:

$$\begin{aligned}\min_{P,\beta,L} \quad & \beta \\ \text{s.t.} \quad & P \succeq 0,\ \beta > 0 \\ & \begin{bmatrix} P & P(A+BLC) & PB & 0 \\ (A+BLC)^T P & P & 0 & C_p^T \\ B^T P & 0 & \beta I & D_p^T \\ 0 & C_p & D_p & \beta I \end{bmatrix} \succeq 0\end{aligned}$$

where $\beta^2 = \gamma$.

The objective of the H_- design problem is to find an observer gain K, so as to increase the effect of the disturbance/attack signal in the detection output. The classical H_- design optimization problem is:

$$\begin{aligned}||\Sigma||_{H_-}^2 \triangleq \max_{K} \quad & \min_{a \in \ell_{2e}} \quad ||y_r||_{\ell_2}^2 \\ & \text{s.t.} \quad ||a||_{\ell_2}^2 = 1, \quad x[0] = 0\end{aligned} \tag{11}$$

where the optimal value of the H_- index is $||\Sigma||_{H_-}$. The BMI approach for the optimization problem (11) exist in the literature (Wang and Yang, 2008) and is given as:

$$\begin{aligned}\max_{P,\gamma,K} \quad & \gamma \\ \text{s.t.} \quad & P \succeq 0,\ \gamma > 0 \\ & \begin{bmatrix} P - C_r^T C_r & -C_r D_r & (A-KC)P \\ -D_r^T C_r^T & -D_r^T D_r + \gamma I & B^T P \\ P(A-KC)^T & PB^T & -P \end{bmatrix} \preceq 0.\end{aligned}$$

Remark 6. The H_∞ design approach reduces the gain of the system Σ_p at a frequency z_1 where the attack impact on the performance output is the highest. The H_- design approach increases the gain of the system Σ_r at a frequency z_2 where the attack impact on the detection output is the lowest. The disadvantage of using these metrics for system design is that, the frequency at which the fault detection occurs might not necessarily be the same frequency at which robustification occurs ($z_1 \neq z_2$).

4. JOINT DESIGN FOR STRICTLY PROPER SYSTEMS

Before introducing the design procedure for solving the non-convex BMIs in the previous section, there are certain limitations faced by these design metrics when applied to certain classes of systems. These limitations are discussed in this section.

Theorem 7. Consider a DT system $\Sigma \triangleq (A, B, C, D)$ which is assumed to be controllable and observable. Let $y_i[k] = C_i x[k] + D_i u[k],\ i = \{1, 2\}, D_1 = 0, D_2 \neq 0$ and $s(\cdot) = \gamma||y_1[k]||_2^2 - ||y_2[k]||_2^2, \gamma \geq 0$. Under these definitions, the system Σ cannot be dissipative with respect to the supply function $s(\cdot)$.

Proof. For the system dynamics defined in the statement of the theorem, using *Proposition 3*, the dissipation inequality (7) becomes:

$$\begin{bmatrix} A^T PA - P & A^T PB \\ B^T PA & B^T PB \end{bmatrix} - \gamma \begin{bmatrix} C_1^T \\ 0 \end{bmatrix} [C_1\ 0] + \begin{bmatrix} C_2^T \\ D_2^T \end{bmatrix} [C_2\ D_2] \preceq 0. \tag{12}$$

For this inequality (12) to be satisfied, the following inequality should be satisfied as well (one of the elements of the above inequality): $B^T PB + D_2^T D_2 \preceq 0$. We know that, with the condition $P \succeq 0$, this inequality is impossible to solve (since $D_2^T D_2 \succ 0$). Hence the LMI (12) is infeasible, which concludes the proof.

Remark 8. For H_- control, in accordance with the definitions of *Theorem 7*, we have $y_1 = y_r$ for which the system matrix D_1 becomes 0 for any system of the form (4). This satisfies the assumption made in the theorem above ($D_1 = 0$). Hence, *Theorem 7* imposes a limitation on the H_- control/design approach.

Remark 9. For H_∞ control, in accordance with the definitions of *Theorem 7*, we have $y_1 = a$ for which the system matrix D_1 becomes I for any system of the form (4). This does not satisfy the assumption made in the theorem above ($D_1 = 0$). Hence, *Theorem 7* does not impose any limitations on the H_∞ control/design approach.

4.1 Revised system description

This limitation described by *Theorem 7* was also a result pointed out in a previous work (see Teixeira (2019)) where an approach to circumvent this issue was proposed based on cyclo-dissipativity which unfortunately leads to a complex design problem. In this section, we will concentrate on an alternative approach. Let us momentarily consider a CT plant described by:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t).\end{aligned}$$

The DT version of the plant is obtained by a zero sample hold (ZOH). The ZOH method can be represented by:

$$y(t = kh) = C \int_{(k-1)h}^{kh} \dot{x}(t)dt + Du(t \in [(k-1)h, kh)) \tag{13}$$

where k is the discrete time step and h represents the sample time.

This effect, mentioned in (13), is not captured in the system description (1)-(3). Hence an system description with the following variation which respects causality (Blachuta, 1999) is adopted.

$$\begin{aligned}y[k] &= Cx_p[k] + D\tilde{u}[k-1] \\ y_p[k] &= C_J x_p[k] + D_J \tilde{u}[k-1]\end{aligned}$$

Defining $e[k] \triangleq x_p[k] - \hat{x}_p[k]$ and $\bar{x}[k] \triangleq [x_p[k]^T\ e[k]^T\ u[k-1]^T\ a[k-1]^T]^T$, the closed-loop system under attack with the performance output and detection output as system outputs becomes:

$$\mathcal{P}_{cl} : \begin{cases} \bar{x}[k+1] &= \bar{A}\bar{x}[k] + \bar{B}a[k] \\ y_p[k] &= \bar{C}_p \bar{x}[k] \\ y_r[k] &= \bar{C}_r \bar{x}[k], \end{cases} \tag{14}$$

where

$$\bar{A} = \begin{bmatrix} A+BLC & 0 & BLD & BLD \\ 0 & A-KC & -KD & -KD \\ LC & 0 & LD & LD \\ 0 & 0 & 0 & 0 \end{bmatrix}, \bar{B} = \begin{bmatrix} B \\ B \\ 0 \\ I \end{bmatrix}$$

$$\bar{C}_p = [C_J\ 0\ D_J\ D_J], \qquad \bar{C}_r = [0\ C\ D\ D].$$

Henceforth in this article, the altered system description (14) will be used unless stated otherwise.

For the altered system description (14), the results discussed in Section 3 still holds which is summarized by the following lemma.

Lemma 10. The *OOG* for any given system of the form (14) is the solution of the optimization problem:

$$||\Sigma||^2_{\ell_{2e},y_p\leftarrow y_r} = \min_{P=P^T\succeq 0,\gamma} \quad \gamma \\ \text{s.t.} \quad \bar{R}(P) \preceq 0,$$

where

$$\bar{R}(P) = \begin{bmatrix} \bar{A}^T P \bar{A} - P & \bar{A}^T P \bar{B} \\ \bar{B}^T P \bar{A} & \bar{B}^T P \bar{B} \end{bmatrix} - \begin{bmatrix} \gamma \bar{C}_r^T \bar{C}_r - \bar{C}_p^T \bar{C}_p & 0 \\ 0 & 0 \end{bmatrix}.$$

4.2 Joint controller and detector design algorithm

In the previous section, an altered system description was proposed so that the *OOG* is bounded for any system of the form (4). With this altered system description, we aim to design an *OOG* optimal controller and detector in this section. This design algorithm is discussed below.

Theorem 11. Consider the closed-loop system described in (14). Let $D = 0$, Define $P \triangleq \begin{bmatrix} P_x & P_{xu} \\ P_{xu}^T & P_u \end{bmatrix}$ where $P_x \in \mathbb{R}^{(2n_x+n_u)\times(2n_x+n_u)}, P_{xu} \in \mathbb{R}^{(2n_x+n_u)\times n_u}, P_u \in \mathbb{R}^{n_u\times n_u}, \tilde{A} \triangleq \begin{bmatrix} A+BLC & 0 & 0 \\ 0 & A-KC & 0 \\ LC & 0 & 0 \end{bmatrix}, \tilde{C}_p \triangleq [C_J \ 0 \ D_J]$ and $\tilde{C}_r \triangleq [0 \ C \ 0]$. The *OOG* optimal controller and detector for the system (14) are obtained by solving the following optimization problem:

$$\begin{aligned} \min_{P=P^T,L,K,\gamma} \quad & \gamma \\ \text{s.t.} \quad & P_x \succ 0 \\ & P\bar{B} = 0 \\ & \begin{bmatrix} -P_x & P_x\tilde{A} & 0 \\ \tilde{A}^T P_x & -P_x & -P_{xu} \\ 0 & -P_{xu}^T & -P_u \end{bmatrix} + \bar{Q}(\gamma) \preceq 0, \end{aligned} \tag{15}$$

where $\bar{Q}(\gamma) \triangleq \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\gamma\tilde{C}_r^T\tilde{C}_r + \tilde{C}_p^T\tilde{C}_p & \tilde{C}_p^T D_J \\ 0 & D_J^T\tilde{C}_p & \tilde{D}_J^T D_J \end{bmatrix}$. The optimal solution variables $L^\star$ and $K^\star$ represent the optimal control and observer gains, where as the optimal value γ^* corresponds to the optimal *OOG*.

Proof. The dissipation inequality for any system of the form (14) can be written as the BMI $\bar{R} \preceq 0$. For the above BMI to be satisfied, the necessary condition must be satisfied as well: $\bar{B}^T P \bar{B} \preceq 0$. Recalling the constraint $P \succeq 0$, as required for dissipativity, the former necessary condition can be rewritten as $P\bar{B} = 0$.

Under the constraint $P\bar{B} = 0$, the BMI $\bar{R} \preceq 0$ reduces to

$$\bar{A}^T P \bar{A} - P + \bar{C}_p^T \bar{C}_p - \gamma \bar{C}_r^T \bar{C}_r \preceq 0 \tag{16}$$

Under the definitions $D = 0, \bar{A} = \begin{bmatrix} \tilde{A} & 0 \\ 0 & 0 \end{bmatrix}, P = \begin{bmatrix} P_x & P_{xu} \\ P_{xu}^T & P_u \end{bmatrix}$ and by using the Schur lemma (Boyd and Vandenberghe, 2004), (16) becomes the last constraint of (15), which concludes the proof.

This optimization problem (15) is non-convex due to the BMI in K, L and P_x. Hence, an alternating minimization approach is proposed in **Algorithm 1** (Li et al., 2019).

Result: K^*, L^*, γ^*, P^*
Initialization: Stabilizing K and $L, k := 1$;
while $||P_k - P_{k-1}|| \geq \epsilon$ **do**
  (i) Solve (15) with respect to P.
  (ii) Update P_k with the result of step (i).
  (iii) Solve (15) with respect to K and L.
  (iv) Update L_k and K_k with the result of step (iii).
end
Algorithm 1: Joint design of controller and detector

The advantages of designing the controller and detector with **Algorithm 1** are: (a) as a direct consequence of a unified objective function, a joint optimal design of K and L is possible, (b) imposing the condition $P_x \succ 0$ guarantees that the system is stable, which is an added advantage (c) if the objective only concerns with the design of controller (or detector), this algorithm still applies but the steps (iii) and (iv) of **Algorithm 1** is an optimization problem w.r.t. L (or K) only.

5. NUMERICAL EXAMPLE

In this section, the effectiveness of the proposed **Algorithm 1** is depicted through a numerical example.

Example 12. Consider the system described in (14) with $A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 0.5 & 0 \\ 0 & 1 & -0.5 \end{bmatrix}, B = \begin{bmatrix} 1 \\ 0.3 \\ 1 \end{bmatrix}, C = 1_{1\times 3}, C_J^T = \begin{bmatrix} I_3 \\ 0_{3\times 1} \end{bmatrix}$, and $D_J = \begin{bmatrix} 0_{3\times 1}^T & 1 \end{bmatrix}^T$. The sub-optimal gains obtained by **Algorithm 1** for the altered system description (14) are $L^* = -0.6208$ and $K^{*T} = [1.2058 \ 0.5283 \ 0.3440]$. For comparison, the classical design problems are solved as follows:

- The H_∞ optimal controller for the altered system description (14) is traditionally computed by solving the BMI (Hilhorst et al., 2014). Since this problem is non-convex, **Algorithm 1** is employed to solve this BMI. The only difference is that the optimization problem is now only a function of L, γ and P.
- When designing the detector using the H_- method for a strictly proper system as in our case, the limitation mentioned in *Remark 8* occurs. This implies that an optimal detector cannot be designed. Hence a sub-optimal detector design for a finite frequency range is adopted (see *Theorem 1*, Wang and Yang (2008)). A frequency range in which a solution exist for this system ($\omega \in [1, 50)$) is chosen.

The design gains obtained from these methods are $L = -0.5787$ and $K^T = [1.6997 \ 0.6185 \ 0.3083]$. The singular values of the altered system with these controller gains are shown in (a) Fig. 1 for performance output ($\bar{\sigma}(\Sigma_p)$) (b) Fig. 2 for detection output ($\underline{\sigma}(\Sigma_r)$). Fig. 1 and Fig. 2 represent the singular values of the system on the unit circle of the complex plane i.e.: the x-axis of the figures represent z where $z = e^{j\omega T_s}$ with $\omega \in [0, \frac{\pi}{T_s}]$. The dark line at $\frac{\pi}{T_s}$ rad/sec represents the Nyquist sampling frequency.

5.1 Discussion

As mentioned before, the objective of the H_∞ design is to minimize the effect of the attack on the performance

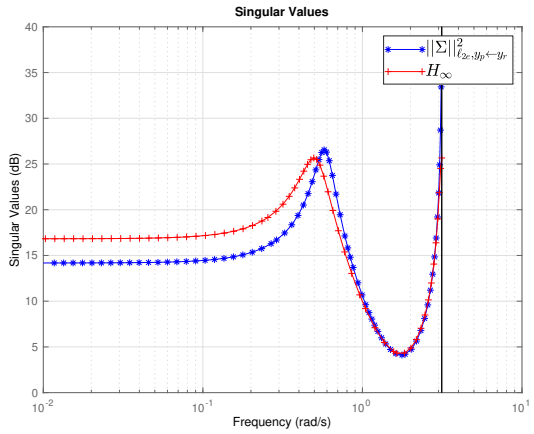

Fig. 1. Singular values - Performance output ($\bar{\sigma}(\Sigma_p)$)

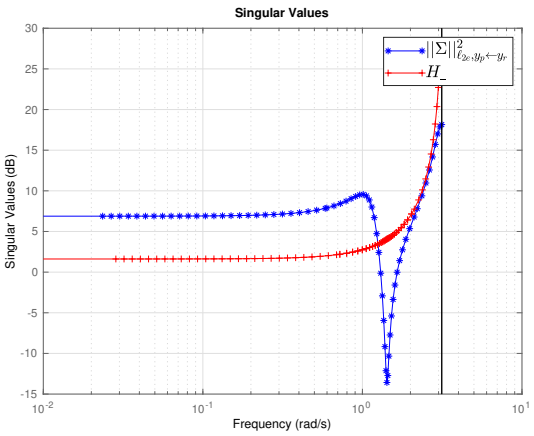

Fig. 2. Singular values - Detection output ($\underline{\sigma}(\Sigma_r)$)

outputs and the aim of the H_- design is to increase the effect of the attack on the detection output. In terms of singular values (SV), these objectives would translate to: aiming to have low SV for the former and high SV for the latter. From Fig. 1, it is clear that in low frequency the SV of our *OOG* method performs as expected and better than the H_∞ design. From Fig. 2, it can be inferred that the detection energy of our *OOG* method is increased at almost all frequency ranges. There is a sharp drop at $\omega \approx 1.5\,\text{rad/s}$. The reasoning behind this is as follows: As the effect of the attack signal in the performance output is negligible, the detector does-not focus on improving the detection at these frequency ranges.

Let us now consider a step attack signal of the form:

$$a[k] = \begin{cases} 1, & k \geq 1 \\ 0, & \text{otherwise.} \end{cases}$$

As shown in Fig. 3, the effect of the attack in the normalized performance output energy ($\frac{1}{k}||y_p||^2_{[0,k]}$) is significantly reduced. In addition, the effect of the attack signal is prominent in the normalized detector output ($\frac{1}{k}||y_r||^2_{[0,k]}$).

The worst-case attack for the performance output is at $\omega = 0.6\,\text{rad/sec}$. Applying this worst-case input signal $a[k] = \sqrt{2}\sin(0.60T_s k)$, the performance of the system

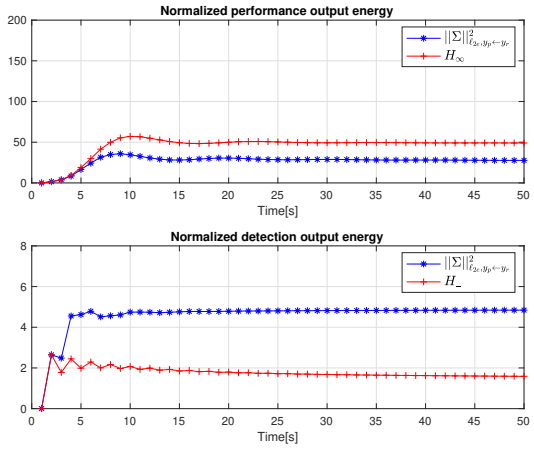

Fig. 3. System outputs for step attack

is shown in Fig. 4. It can be seen that the detection and the performance energy is higher than the classical metrics. Hence, this method can help to improve attack detection, at the expense of an increased deterioration of the performance. However, in practice, the performance deterioration may be prevented by timely switching to a fault-tolerant controller (Gao, 2015) when the attack is detected.

The worst-case attack for the detection output is at $\omega = 1.42$ rad/sec. Applying this worst-case input signal $a[k] = \sqrt{2}\sin(1.42T_s k)$, the performance of the system is shown in Fig. 5. Although the detection energy is lower than the classical metrics, it can be noted that the detection energy at time intervals $[0, 5]$ s is very close to that of the classical metric which can help in detection. Moreover, although the detection is harder in this case, the performance degradation is also less significant when compared to the other attack signals examined in this section. By shifting the focus from attacks with small impact on performance, the detector has managed to improve its detection capabilities for other attacks with higher impact.

With the H_∞ and H_- optimal controller and detector obtained from **Algorithm 1**, we obtain $\gamma = 238.4265$ by solving (8). With the *OOG* optimal controller and detector parameters from **Algorithm 1**, we obtain $\gamma = 74.7920$ by solving (8). This is an indication of the bound imposed by the choice of the design metric chosen on the disruption induced by the attack signal.

6. CONCLUSION

This paper considers actuator attacks that aim at maximizing impact while remaining undetected, and proposes a joint detector and controller design approach based on dissipative systems theory and alternating minimization. Numerical examples are provided to compare the effectiveness of the proposed approach. Future work directions include the extension to other classes of attacks, as well as to other controller structures.

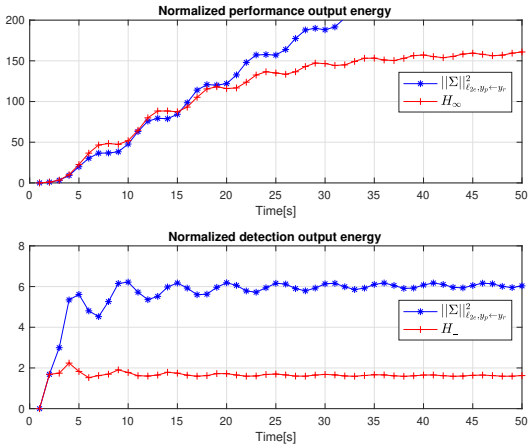

Fig. 4. System outputs for $a[k] = \sqrt{2}\sin(0.60T_s k)$

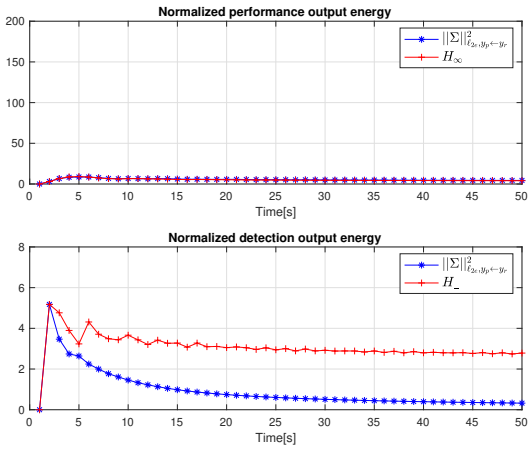

Fig. 5. System outputs for $a[k] = \sqrt{2}\sin(1.42T_s k)$

REFERENCES

Bishop, M.A. (2002). The art and science of computer security.

Blachuta, M.J. (1999). Discrete-time modeling of sampled-data control systems with direct feedthrough. *IEEE Trans. Autom. Control*, 44(1), 134–139.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

Cárdenas, A.A., Amin, S., Lin, Z.S., Huang, Y.L., Huang, C.Y., and Sastry, S. (2011). Attacks against process control systems: risk assessment, detection, and response. In *ACM symposium on information, computer and commun. security*, 355–366.

Chong, M.S., Sandberg, H., and Teixeira, A.M. (2019). A tutorial introduction to security and privacy for cyber-physical systems. In *2019 18th European Control Conf. (ECC)*, 968–978. IEEE.

Ding, S., Zhong, M., Jeinsch, T., and Tang, B. (2002). LMI-based integration of robust H_∞-control and RFD for LTI systems. *IFAC Proceedings Volumes*, 35(1), 161–166.

Fawzi, H., Tabuada, P., and Diggavi, S. (2014). Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Trans. Autom. Control*, 59(6), 1454–1467.

Ferrari, R.M. and Teixeira, A.M. (2017). Detection and isolation of routing attacks through sensor watermarking. In *2017 American Control Conf. (ACC)*, 5436–5442. IEEE.

Gao, Z. (2015). Fault estimation and fault-tolerant control for discrete-time dynamic systems. *IEEE Trans. Ind. Electron*, 62(6), 3874–3884.

Hilhorst, G., Pipeleers, G., Oliveira, R.C., Peres, P.L., and Swevers, J. (2014). On extended LMI conditions for H_2/H_∞ control of discrete-time linear systems. *IFAC Proceedings Volumes*, 47(3), 9307–9312.

Langner, R. (2011). Stuxnet: Dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3), 49–51.

Li, Q., Zhu, Z., and Tang, G. (2019). Alternating minimizations converge to second-order optimal solutions. In *Int. Conf. Machine Learning*, 3935–3943.

Milošević, J., Sandberg, H., and Johansson, K.H. (2018a). A security index for actuators based on perfect undetectability: Properties and approximation. In *2018 56th Annual Allerton Conf. on Commun., Control, and Comput.*, 235–241. IEEE.

Milošević, J., Umsonst, D., Sandberg, H., and Johansson, K.H. (2018b). Quantifying the impact of cyber-attack strategies for control systems equipped with an anomaly detector. In *2018 European Control Conf. (ECC)*, 331–337. IEEE.

Mo, Y., Weerakkody, S., and Sinopoli, B. (2015). Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Systems Magazine*, 35(1), 93–109.

Pasqualetti, F., Dorfler, F., and Bullo, F. (2015). Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1), 110–127.

Tan, D. and Patton, R.J. (2015). Integrated fault estimation and fault tolerant control: A joint design. *IFAC-PapersOnLine*, 48(21), 517–522.

Teixeira, A., Sandberg, H., and Johansson, K.H. (2015). Strategic stealthy attacks: the output-to-output l_2-gain. In *2015 54th IEEE Conf. on Decision and Control (CDC)*, 2582–2587. IEEE.

Teixeira, A., Shames, I., Sandberg, H., and Johansson, K.H. (2012). Revealing stealthy attacks in control systems. In *2012 50th Annual Allerton Conf. on Commun., Control, and Computing*, 1806–1813. IEEE.

Teixeira, A.M. (2019). Optimal stealthy attacks on actuators for strictly proper systems. In *2019 IEEE 58th Conf. on Decision and Control (CDC)*, 4385–4390. IEEE.

Urbina, D.I., Giraldo, J.A., Cardenas, A.A., Tippenhauer, N.O., Valente, J., Faisal, M., Ruths, J., Candell, R., and Sandberg, H. (2016). Limiting the impact of stealthy attacks on industrial control systems. In *ACM SIGSAC Conf. on Computer and Commun. Security*, 1092–1105.

Wang, H. and Yang, G.H. (2008). A finite frequency domain approach to fault detection for linear discrete-time systems. *Int. Journal of Control*, 81(7), 1162–1171.

Ye, D. and Luo, S. (2019). A co-design methodology for cyber-physical systems under actuator fault and cyber attack. *Journal of the Franklin Institute*, 356(4), 1856–1879.