Author/s:
Murguia, C;Shames, I;Ruths, J;Nešić, D

Title:
Security metrics and synthesis of secure control systems

Date:
2020-05-01

Citation:
Murguia, C., Shames, I., Ruths, J. & Nešić, D. (2020). Security metrics and synthesis of secure control systems. Automatica, 115, https://doi.org/10.1016/j.automatica.2019.108757.

Persistent Link:
https://hdl.handle.net/11343/297904

Security Metrics and Synthesis of Network Control Systems (extended preprint) ⋆

Carlos Murguia [a], Iman Shames [a], Justin Ruths [b], Dragan Nešić [a]

[a] *Department of Electrical and Electronic Engineering, University of Melbourne, Australia*

[b] *Departments of Mechanical and Systems Engineering, University of Texas at Dallas, USA*

Abstract

As more attention is paid to security in the context of control systems and as attacks occur to real control systems throughout the world, it has become clear that some of the most nefarious attacks are those that evade detection. The term *stealthy* has come to encompass a variety of techniques that attackers can employ to avoid being detected. In this manuscript, for a class of perturbed linear time-invariant systems, we propose two *security metrics* to quantify the potential impact that *stealthy attacks* could have on the system dynamics by tampering with sensor measurements. We provide analysis mathematical tools (in terms of linear matrix inequalities) to quantify these metrics for given system dynamics, control structure, system monitor, and set of sensors being attacked. Then, we provide synthesis tools (in terms of semidefinite programs) to redesign controllers and monitors such that the impact of stealthy attacks is minimized and the required attack-free system performance is guaranteed.

Key words: Network Control Systems; Model-based fault/attack monitors, Security Metrics, Secure Control, Attacks.

1 Introduction

Recently, there has been significant interest and work in the broad area of security of Networked Control Systems (NCSs), see, e.g., [1,6,18,20,23,25–27,29,35]. Security of NCSs investigates properties of conventional control systems in the presence of adversarial disturbances. Control theory has shown great ability to robustly deal with disturbances and uncertainties. However, adversarial attacks raise all-new issues due to the aggressive and strategic nature of the disturbances that attackers might inject into the system.

This paper focuses on quantifying and minimizing attacker capabilities in NCSs. A majority of the work on attack detection leverages the established literature of fault detection [6,8,19,29]. Fault detection techniques use an *estimator* to forecast the evolution of the system dynamics. When the residual (the difference between measurements and their estimates) is larger than a predetermined threshold, an alarm is raised. Fault detectors impose limits on attacks if the attacker aims at avoiding being identified. Beyond retooling existing fault detection techniques for the new attack detection context, a fundamental question is: given a fault detection scheme, how does this particular scheme constrain the influence of an attacker? More specifically, what is an attacker able to accomplish when the system employs certain fault detection procedure?

Different methodologies exist for evaluating the impact of attacks. Most of the existing work uses some measure of state deviation. A number of groups have studied the system response when attacks are constrained by a fault detector, i.e., they look for the system trajectories that can be induced due to *stealthy attacks* – attacks such that the detector threshold is never crossed [7,11,25,28,29]. In this manuscript, we provide mathematical tools for *quantifying* and *minimizing* the potential impact of sensor stealthy attacks on the system dynamics. We consider the set of states that stealthy attacks can induce in the system (*the attacker's stealthy reachable set*) and use the "size" of this set, in terms of volume (Lebesgue measure), as a *security metric* for NCSs. Stealthy reachable sets provide insight of the system potential performance degradation induced by stealthy attacks. The volume of these sets quantifies the size of the portion of the state space that opponents could reach by tampering with particular subsets of sen-

⋆ This paper was not presented at any IFAC meeting. Corresponding author Carlos Murguia.

Email addresses: carlos.murguia@unimelb.edu.au (Carlos Murguia), iman.shames@unimelb.edu.au (Iman Shames), jruths@utdallas.edu (Justin Ruths), dnesic@unimelb.edu.au (Dragan Nešić).

sors. So, for different subsets of sensors being attacked, we have reachable sets of different sizes; and thus, one could quantify the system sensitivity to attacks in particular sensors by comparing their corresponding volumes. Because it is not mathematically tractable to compute stealthy reachable sets exactly, we provide analysis tools – in terms of Linear Matrix Inequalities (LMIs) – for computing *ellipsoidal outer approximations* of the attacker's reachable sets. The obtained approximations quantify the attacker's potential impact when it is constrained to stay hidden from the detector. We use the size (again in terms of volume) of these ellipsoidal approximations to approximate the proposed security metric (the volume of stealthy reachable sets). Note that this security metric just tell us which sensors lead to larger reachable sets. So, we are implicitly assuming that all points in the state space are equally important (in terms of security) and we are just interested in the overall "number" of states potentially reachable by attacks. However, if there are regions of the state space that are more important than others (again in terms of security), and thus we are particularly interested in knowing if these regions are reachable by attacks, we need a different metric. To this end, as a second security metric, we propose the minimum distance from the attacker's reachable set to a possible set of *critical states* – states that, if reached, compromise the integrity or safe operation of the system. We approximate this distance by the minimum distance between the ellipsoidal approximations and the critical states. This distance gives us intuition on how far the actual attacker's reachable set is from the critical states. Once we have provided a complete set of analysis tools to approximate the aforementioned security metrics, we use these tools to derive synthesis tools (in terms of semidefinite programs) to redesign controllers and fault detectors such that the impact of stealthy attacks is minimized and the required attack-free system performance is guaranteed.

There are a few results in this direction already; chiefly the work in [24] (and the preliminary paper [23]), where the authors provide a recursive algorithm to compute ellipsoidal approximations of attacker's reachable sets for Linear Time Invariant (LTI) systems subjected to Gaussian noise. The authors in [24] give *analysis-only* results for a very particular structure of controllers and fault-detectors. They consider Kalman-filter based fault detectors and use the state of the filter to construct output feedback controllers. Although this results in compact designs of controllers and fault detectors, the flexibility of having dedicated controllers and detectors (mainly for synthesis of secure control systems) is limited. We remark that, in the stochastic setting considered in [24], the detector threshold is always crossed even when there are no attacks. This is due to the infinite support of the Gaussian noise they consider. Thus, they do not consider stealthy attacks in the sense described above. Instead, they consider attacks that increase the alarm rate of the detector by a small amount only. Then, they approximate the attacker's reachable set corresponding to this small increase.

The *main contributions* of this manuscript (in contrast to the work in [24] and in general) are the following: 1) we provide a set of mathematical tools *in terms of semidefinite programs* to approximate reachable sets induced by *stealthy attacks* for LTI systems driven by *peak bounded deterministic perturbations*; 2) we provide both *analysis and synthesis* results for *dedicated* general dynamic output feedback controllers and observer-based fault detectors; 3) we propose *two security metrics* to assess the vulnerability of systems to attacks, and *optimize these metrics* (enhancing thus the system resilience to attacks) by synthesizing optimal controllers and detectors; and 4) the synthesis part considers the *attack-free performance of the closed-loop dynamics*, i.e., we optimize the security metrics subject to certain prescribed attack-free system performance. In our preliminary work [28], we also approximate reachable sets of false-data-injection attacks but we consider the same stochastic framework as the one proposed in [24], i.e., Gaussian noise, joint Kalman-filter based fault detectors and controllers, and attacks increasing the alarm rate of the detector. Thus, the problems considered in this manuscript (and the obtained results) and the ones addressed in [28] are fundamentally different; and the set of results (and the tools used to obtain them) are different too. Moreover, in [28], we consider attacks to all the sensors. Although the latter case provides a worse-case scenario, we lose the capability of quantifying the sensitivity of the system dynamics to attacks on specific sensors. As in [24], the results in [28] mainly focus on analysis (although they hint how to address synthesis for joint Kalman-filter based detectors and controllers).

There are a few other results that considers different security metrics for control systems. However, all of those results have a very different interpretation and the framework upon which they are constructed is fundamentally very different from ours. For instance, in [1,2], for arbitrary detection procedures, the authors quantify how much the attacker can increase the asymptotic covariance (their security metric) of state estimates while remaining stealthy. They characterize stealthiness using the *Kullback-Leibler Divergence* [30] between the attack-free and the attacked estimates. In [34,36], the authors use the notion of *security index* for LTI systems. This index refers to the smallest number of sensors and actuators that have to be compromised for successfully launching stealthy attacks. For linear stochastic systems, the authors is [21] propose two security metrics: the probability that some of the critical states leave a safety region; and the expected value of the infinity norm of the critical states. Finally, in [22], tools from *finance risk theory* are used to quantify security of LTI systems.

The remainder of the paper is organized as follows. In Section 2, we present some preliminaries results needed for the subsequent sections. We provide tools for computing outer time-varying bounds on the trajectories

of a class of perturbed nonlinear discrete-time systems. Then, we use these tools to obtain outer ellipsoidal approximations of reachable sets of LTI systems driven by multiple peak bounded perturbations. The system dynamics, monitor, and controller descriptions are given in Section 3. Our proposed security metrics and analysis tools, together with some numerical results, are given in Section 4; and the corresponding synthesis results are given in Section 5. Finally, conclusions and recommendations are stated in Section 6.

2 Preliminaries

In this section, we present some preliminary results needed for the subsequent sections. First, in Lemma 1, we present a preliminary tool used to compute outer time-varying bounds on the trajectories of perturbed discrete-time systems. Next, in Proposition 1, we use this lemma to compute outer ellipsoidal approximations of reachable sets of LTI systems driven by multiple peak bounded perturbations.

Lemma 1 *For a given $a \in (0,1)$, if there exist functions $a_k^i : \mathbb{N} \to (0,1)$, $i = 1,\dots,N$, and $V : \mathbb{R}^{n_\xi} \to \mathbb{R}_{\geq 0}$ satisfying $\sum_{i=1}^N a_k^i \geq a$ and, for all $k \in \mathbb{N}$, the inequality:*

$$V(\xi_{k+1}) - aV(\xi_k) - \sum_{i=1}^{N}(1-a_k^i)(\omega_k^i)^T W_k^i \omega_k^i \leq 0; \quad (1)$$

then, $V(\xi_k) \leq \alpha_k$, where $\alpha_k := a^{k-1}V(\xi_1) + \frac{(N-a)(1-a^{k-1})}{1-a}$, and $\lim_{k\to\infty} V(\zeta_k) \leq \frac{N-a}{1-a}$.

Proof: By assumption, $(\omega_k^i)^T W_k^i \omega_k^i \leq 1$, for $i = 1,\dots,N$; then, from (1), we have

$$\begin{aligned} V(\xi_{k+1}) &\leq aV(\xi_k) + \sum_{i=1}^{N}(1-a_k^i)\underbrace{(\omega_k^i)^T W_k^i \omega_k^i}_{\leq 1} \\ &\leq aV(\xi_k) + (N-a), \quad (2) \end{aligned}$$

because $\sum_{i=1}^N a_k^i \geq a$. It follows that

$$V(\xi_k) \leq aV(\xi_{k-1}) + (N-a), \quad (3)$$
$$V(\xi_{k-1}) \leq aV(\xi_{k-2}) + (N-a). \quad (4)$$

Using (4) to upper bound (3) and continuing the recursion yields

$$V(\xi_k) \leq a^{k-1}V(\xi_1) + \frac{(N-a)(1-a^{k-1})}{1-a}.$$

Therefore, $\lim_{k\to\infty} V(\zeta_k) \leq (N-a)/(1-a)$ because $a \in (0,1)$. ■

Next, we present a tool to identify outer ellipsoidal approximations of reachable sets of LTI systems driven by multiple peak bounded perturbations.

Consider the perturbed LTI system

$$\xi_{k+1} = A\xi_k + \sum_{i=1}^{N} B^i \omega_k^i, \quad (5)$$

with $k \in \mathbb{N}$, state $\xi_k \in \mathbb{R}^{n_\xi}$, initial condition $\xi_1 \in \mathbb{R}^{n_\xi}$, perturbation $\omega_k^i \in \mathbb{R}^{p_i}$ satisfying $(\omega_k^i)^T W_i \omega_k^i \leq 1$ for some positive definite matrix $W_i \in \mathbb{R}^{p_i \times p_i}$, $i = 1,\dots,N$, $N \in \mathbb{N}$, and matrices $A \in \mathbb{R}^{n_\xi \times n_\xi}$ and $B^i \in \mathbb{R}^{n_\xi \times p_i}$. Denote by $\psi^\xi(k, \xi_1, \omega^1(\cdot), \cdots, \omega^N(\cdot)) := A^{k-1}\xi_1 + \sum_{i=1}^N \sum_{j=0}^{k-2} A^j B^i \omega_{k-1-j}^i$ the solution of (5) at time instant $k > 1$ given the initial condition ξ_1 and the infinite disturbance sequence $\omega^i(\cdot) := \{\omega_1^i, \omega_2^i, \dots\}$.

Definition 1 *The reachable set $\mathcal{R}_k^\xi$ at time instant $k > 1$ from initial condition ξ_1, is the set of states $\psi^\xi(k, \xi_1, \omega^1(\cdot), \cdots, \omega^N(\cdot))$ reachable in k steps by system (5) through all possible perturbations satisfying $(\omega_k^i)^T W_i \omega_k^i \leq 1$, i.e.,*

$$\mathcal{R}_k^\xi := \left\{ \xi \in \mathbb{R}^{n_\xi} \;\middle|\; \begin{array}{l} \xi = \psi^\xi(k, \xi_1, \omega^1(\cdot), \cdots, \omega^N(\cdot)), \\ \xi_1 \in \mathbb{R}^{n_\xi}, \text{ and } (\omega_k^i)^T W_i \omega_k^i \leq 1. \end{array} \right\}.$$

Proposition 1 *Consider the LTI system (5) and the reachable set $\mathcal{R}_k^\xi$ introduced in Definition 1. For a given $a \in (0,1)$, if there exist constants $a_1 = \tilde{a}_1, \dots, a_N = \tilde{a}_N$ and matrix $\mathcal{P} = \tilde{\mathcal{P}} \in \mathbb{R}^{n_\xi \times n_\xi}$ satisfying:*

$$\begin{cases} a_1,\dots,a_N \in (0,1),\ a_1 + \cdots + a_N \geq a, \\ \mathcal{P} > 0,\ \begin{bmatrix} a\mathcal{P} & A^T\mathcal{P} & \mathbf{0} \\ \mathcal{P}A & \mathcal{P} & \mathcal{P}B \\ \mathbf{0} & B^T\mathcal{P} & W_{a_i} \end{bmatrix} \geq 0; \end{cases} \quad (6)$$

with $W_{a_i} := \mathrm{diag}[(1-a_1)W_1, \dots, (1-a_N)W_N] \in \mathbb{R}^{\bar{p}\times\bar{p}}$, $B := (B^1, \dots, B^N) \in \mathbb{R}^{n_\xi \times \bar{p}}$, and $\bar{p} = \sum_{i=1}^N p_i$; then, $\mathcal{R}_k^\xi \subseteq \tilde{\mathcal{E}}_k^\xi := \{\xi \in \mathbb{R}^{n_\xi} \mid \xi^T\tilde{\mathcal{P}}\xi \leq \tilde{\alpha}_k^\xi\}$, where $\tilde{\alpha}_k^\xi := a^{k-1}\xi_1^T\tilde{\mathcal{P}}\xi_1 + \big((N-a)(1-a^{k-1})\big)/(1-a)$.

Proof: For a positive definite matrix $\mathcal{P} \in \mathbb{R}^{n_\xi \times n_\xi}$, let $V_k = \xi_k^T\mathcal{P}\xi_k$ in Lemma 1. Substituting this V_k, the dynamics $\xi_{k+1} = A\xi_k + B\omega_k$, with stacked vector of perturbations $\omega_k := ((\omega_k^1)^T, \dots, (\omega_k^N)^T)^T$, and the inequality $a_1 + \cdots + a_N \geq a$ in (1) yields

$$\nu_k^T \underbrace{\begin{bmatrix} \mathcal{P} - A^T\mathcal{P}A & -A^T\mathcal{P}B \\ -B^T\mathcal{P}A & W_{a_i} - B^T\mathcal{P}B \end{bmatrix}}_{Q} \nu_k \geq 0,$$

with $\nu_k := \left(\xi_k^T, \omega_k^T\right)^T$. This inequality is satisfied if and only if Q is positive semidefinite. This Q can be written as the Schur complement of a higher dimensional matrix Q'; it follows that $Q \geq \mathbf{0} \Leftrightarrow Q' \geq 0$ [4], where

$$Q' := \begin{bmatrix} \mathcal{P} & \mathbf{0} & A^T\mathcal{P} \\ \mathbf{0} & W_{a_i} & B^T\mathcal{P} \\ \mathcal{P}A & \mathcal{P}B & \mathcal{P} \end{bmatrix}.$$

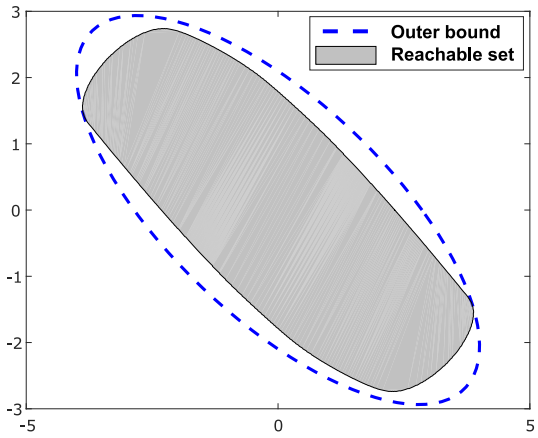

Fig. 1. Tightness of the ellipsoidal outer approximations of reachable sets obtained by Corollary 1.

Consider the congruence transformation $Q' \to \mathcal{T}^T Q' \mathcal{T}$,

$$\mathcal{T} := \begin{bmatrix} I & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \\ \mathbf{0} & I & \mathbf{0} \end{bmatrix}.$$

Hence, $Q \geq \mathbf{0} \Leftrightarrow Q' \geq \mathbf{0} \Leftrightarrow \mathcal{T}^T Q' \mathcal{T} \geq \mathbf{0}$, see [4] for details. Inequality $\mathcal{T}^T Q' \mathcal{T} \geq \mathbf{0}$ equals the last inequality in (6). Then, by Lemma 1, we have $\xi_k^T \tilde{\mathcal{P}} \xi_k \leq a^{k-1} \xi_1^T \tilde{\mathcal{P}} \xi_1 + \big((N-a)(1-a^{k-1})\big)/(1-a) = \tilde{\alpha}_k^\xi$ for any $a_i = \tilde{a}_i$, $i = 1, \ldots, m$, and $\mathcal{P} = \tilde{\mathcal{P}}$ satisfying (6). It follows that the trajectories ξ_k generated by $\xi_{k+1} = A\xi_k + \sum_{i=1}^N B^i \omega_k^i$, the initial condition ξ_1, and the perturbation ω_k, are always contained in the time-varying ellipsoid $\tilde{\mathcal{E}}_k^\xi$, i.e., $\mathcal{R}_k^\xi \subseteq \tilde{\mathcal{E}}_k^\xi$. ■

Remark 1 *Note that the contribution of the initial condition ξ_1 to the sequence $\tilde{\alpha}_k^\xi$ vanishes exponentially. We have that $\lim_{k\to\infty} \tilde{\alpha}_k^\xi = (N-a)/(1-a)$; therefore*

$$\lim_{k\to\infty} \tilde{\mathcal{E}}_k^\xi = \{\xi \in \mathbb{R}^{n_\xi} \mid \xi^T \tilde{\mathcal{P}} \xi \leq (N-a)/(1-a)\} =: \tilde{\mathcal{E}}_\infty^\xi. \tag{7}$$

That is, after transients due to initial conditions have settled down, all trajectories of the system are trapped inside the ellipsoid $\tilde{\mathcal{E}}_\infty^\xi$.

Proposition 1 provides a tool for computing time-varying ellipsoidal outer approximations $\tilde{\mathcal{E}}_k^\xi$ of $\mathcal{R}_k^\xi$. Note that $\tilde{\mathcal{E}}_k^\xi$ could be an arbitrarily conservative approximation of $\mathcal{R}_k^\xi$ as long as $\mathcal{R}_k^\xi \subseteq \tilde{\mathcal{E}}_k^\xi$. Then, to make $\tilde{\mathcal{E}}_k^\xi$ less conservative, we aim at obtaining ellipsoids with *minimal volume*, i.e., the tightest possible ellipsoid bounding $\mathcal{R}_k^\xi$ among all the ellipsoids generated by Proposition 1. To find such an ellipsoid, we look to minimize $(\det[\mathcal{P}])^{-1/2}$ subject to (6) because $(\det[\mathcal{P}])^{-1/2}$ is proportional to the volume of the asymptotic ellipsoid $\xi^T \mathcal{P} \xi = (N-a)/(1-a)$ for any $N \in \mathbb{N}$ and $a \in (0,1)$ [4]. We minimize $\log\det[\mathcal{P}^{-1}]$ instead as it shares the same minimizer with $(\det[\mathcal{P}])^{-1/2}$ and because for positive definite $\mathcal{P}$ this objective is convex [4]. This is stated in the following corollary of Proposition 1.

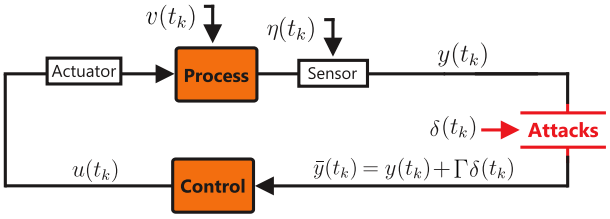

Fig. 2. Cyber-physical system under sensor attacks.

Corollary 1 *Consider the perturbed LTI system (5) and the reachable set $\mathcal{R}_k^\xi$ introduced in Definition 1. For a given $a \in (0,1)$, if there exist constants $a_1 = a_1^*, \ldots, a_N = a_N^*$ and matrix $\mathcal{P} = \mathcal{P}^*$ solution of the convex optimization:*

$$\begin{cases} \min\limits_{\mathcal{P}, a_1, \ldots, a_N} & -\log\det[\mathcal{P}], \\ \text{s.t.} & (6); \end{cases} \tag{8}$$

then, $\mathcal{R}_k^\xi \subseteq \mathcal{E}_k^\xi := \{\xi \in \mathbb{R}^{n_\xi} \mid \xi^T \mathcal{P}^* \xi \leq \alpha_k^\xi\}$, where $\alpha_k^\xi := a^{k-1} \xi_1^T \mathcal{P}^* \xi_1 + \big((N-a)(1-a^{k-1})\big)/(1-a)$. Moreover, for any $a_i = \tilde{a}_i \neq a_i^*$ and $\mathcal{P} = \tilde{\mathcal{P}} \neq \mathcal{P}^*$ satisfying the constraints in (6) and corresponding ellipsoidal approximation $\tilde{\mathcal{E}}_k^\xi$, the volume of $\mathcal{E}_\infty^\xi$ (see (7)) is strictly less than the volume of $\tilde{\mathcal{E}}_\infty^\xi$, i.e., $\mathcal{E}_k^\xi$ has the minimum asymptotic volume among all the outer ellipsoidal approximations $\tilde{\mathcal{E}}_k^\xi$ generated by Proposition 1.

Proof: The solution space of the objective function is convex because the constraints are linear [5]. Moreover, the function $\log\det[\mathcal{P}^{-1}]$ is convex for any positive definite matrix $\mathcal{P}$ [4]. Hence, Corollary 1 follows from Proposition 1, convexity of the solution space, and convexity of the objective function. ■

Remark 2 *Note that the constant $a \in (0,1)$ in Corollary 1 must be fixed before solving (8). This constant is, in fact, a variable of the optimization problem. However, to convexify the cost and linearize some of the constraints, we fix its value before solving (8) and search over $a \in (0,1)$ to find the optimal $\mathcal{P}^*$. The latter increases the computations needed to find $\mathcal{P}^*$; however, because $a \in (0,1)$ (a bounded set), the required grid is of reasonable size. Indeed, we are interested in selecting the $a \in (0,1)$ that leads to the asymptotic ellipsoid with minimum volume.*

In Figure 1, we illustrate the potential tightness of the ellipsoidal outer approximations obtained using Corollary 1. The solid gray area is the actual reachable set obtained by extensive Monte Carlo simulations, and the ellipsoidal approximation is depicted with dashed lines. This figure corresponds to an LTI system driven by two peak bounded perturbation. The exact numerical values of the system matrices and the perturbations' bounds can be found in [14, Section 4].

3 System & Monitor Description

In this section, we introduce the class of systems under study, the monitor that we use to pinpoint attacks, and the control scheme.

3.1 System Dynamics

Consider the LTI perturbed system

$$\begin{cases} x^p(t_{k+1}) = A^p x^p(t_k) + B^p u(t_k) + E v(t_k), \\ y(t_k) = C^p x^p(t_k) + F\eta(t_k), \end{cases} \tag{9}$$

with sampling time-instants $t_k, k \in \mathbb{N}$, state $x^p \in \mathbb{R}^n$, output $y \in \mathbb{R}^m$, control input $u \in \mathbb{R}^l$, matrices A^p, B^p, C^p, E, and F of appropriate dimensions, and unknown system and sensor perturbations $v \in \mathbb{R}^q$ and $\eta \in \mathbb{R}^m$, respectively. The perturbations are assumed to be peak bounded, i.e., $v_k^T v_k \leq \bar{v}$ and $\eta_k^T \eta_k \leq \bar{\eta}$ for some known $\bar{v}, \bar{\eta} \in \mathbb{R}_{>0}$ and all $k \in \mathbb{N}$. The pair (A^p, B^p) is stabilizable and (A^p, C^p) is detectable. At the time-instants $t_k, k \in \mathbb{N}$, the output of the process $y(t_k)$ is sampled and transmitted over a communication network. The received output $\bar{y}(t_k)$ is used to compute control actions $u(t_k)$ which are sent back to the actuators. The complete control-loop is assumed to be performed instantaneously, i.e., sampling, transmission, and arrival time-instants are equal. In this manuscript, we focus on *false data injection attacks* on sensor measurements. That is, in between transmission and reception of sensor data, an attacker may inject data to signals coming from sensors to the controller, see Fig. 2. The opponent compromises up to s sensors, $s \in \{1, \ldots, m\}$ of the system. Denote the attacker's sensor selection matrix $\Gamma \in \mathbb{R}^{m \times s}$, $\Gamma \subseteq \{\gamma_1, \ldots, \gamma_m\}$ where $\gamma_i \in \mathbb{R}^{m \times 1}$ denotes the i-th vector of the canonical basis of $\mathbb{R}^m$. After each transmission and reception, the networked output $\bar{y}$ takes the form:

$$\bar{y}(t_k) := y(t_k) + \Gamma\delta(t_k), \tag{10}$$

where $\delta(t_k) \in \mathbb{R}^s$ denotes *additive sensor attacks/faults*. Denote $x_k := x(t_k)$, $u_k := u(t_k)$, $v_k := v(t_k)$, $\bar{y}_k := \bar{y}(t_k)$, $\eta_k := \eta(t_k)$, and $\delta_k := \delta(t_k)$. Using this new notation, the attacked system is written in the following compact form:

$$\begin{cases} x^p_{k+1} = A^p x^p_k + B^p u_k + E v_k, \\ \bar{y}_k = C^p x^p_k + F\eta_k + \Gamma\delta_k. \end{cases} \tag{11}$$

3.2 Filter and Residual

In this manuscript, we aim at characterizing the effect that false data injection attacks can induce in the system without being detected by standard *fault-detectors*. The main idea behind fault detection is the use of an estimator to forecast the evolution of the system state. If the difference between what it is measured and the output estimation is larger than expected, there may be a fault in or an attack on the system. Here, to estimate the state of the process, we use the filter:

$$\hat{x}_{k+1} = A^p \hat{x}_k + B^p u_k + L\big(\bar{y}_k - C^p \hat{x}_k\big), \tag{12}$$

with estimated state $\hat{x} \in \mathbb{R}^n$ and filter gain matrix $L \in \mathbb{R}^{n \times m}$. Define the estimation error $e_k := x^p_k - \hat{x}_k$. Given the system dynamics (11) and the filter (12), the estimation error dynamics is given by

$$e_{k+1} = \big(A^p - LC^p\big)e_k - L\Gamma\delta_k - LF\eta_k + Ev_k. \tag{13}$$

The pair (A^p, C^p) is detectable; hence, the observer gain L can be selected such that $(A^p - LC^p)$ is Schur. We assume that L is such that $(A^p - LC^p)$ is Schur. Define the *residual* $r_k \in \mathbb{R}^m$

$$r_k := \bar{y}_k - C^p \hat{x}_k = C^p e_k + \Gamma\delta_k + F\eta_k, \tag{14}$$

which evolves according to the difference equation:

$$\begin{cases} e_{k+1} = \big(A^p - LC^p\big)e_k - L\Gamma\delta_k - LF\eta_k + Ev_k, \\ r_k = C^p e_k + \Gamma\delta_k + F\eta_k. \end{cases} \tag{15}$$

3.3 Distance Measure, Anomaly Detection, and System Monitor

The input to any detection procedure is a *distance measure* $z_k \in \mathbb{R}$, i.e., a measure of how deviated the estimator is from the attack-free system dynamics [12]. Here, we use a quadratic form of the residual as distance measure. Consider the residual sequence r_k and some positive definite matrix $\Pi \in \mathbb{R}^{m \times m}$. Define the distance measure $z_k := r_k^T \Pi r_k$ and consider the following monitor.

System Monitor:

$$\text{If } \; z_k = r_k^T \Pi r_k > 1, \;\; \tilde{k} = k. \tag{16}$$

Design parameter: positive semidefinite $\Pi \in \mathbb{R}^{m \times m}$.
Output: alarm time(s) $\tilde{k}$.

Thus, the monitor is designed so that alarms are triggered if z_k exceeds one. The matrix Π must be selected such that, after sufficiently large number of time-steps (enough to allow transients to settle down), $z_k \leq 1$ in the attack-free case. That is, after transients due to initial conditions have decreased to a desired level, the ellipsoid $r_k^T \Pi r_k = 1$ must contain all the possible trajectories that the perturbations v_k and η_k can induce in the residual given Eq. (15) and the inequalities $v_k^T v_k \leq \bar{v}$ and $\eta_k^T \eta_k \leq \bar{\eta}$. Note that the tighter the ellipsoidal bound, the less opportunity the attacker has to manipulate the system without being detected. Here, we use Corollary 1 to design an optimal matrix Π (in terms of tightness of the ellipsoidal bound). In particular, using Corollary 1, we obtain an outer time-varying ellipsoidal approximation of the reachable set of the estimation error (13) driven by v_k and η_k in the attack-free case ($\delta_k = \mathbf{0}$). Once we have this ellipsoid, using the $\mathcal{S}$-procedure [4], we project it onto the residual hyperplane to get the ellip-

soid $r_k^T \Pi r_k = 1$ of the monitor. For transparency, these results are presented in the appendix. We need, however, the following assumption for the subsequent sections.

Assumption 1 *In the attack-free case ($\delta_k = \mathbf{0}$), there exists some $k^* \in \mathbb{N}$ such that the matrix Π of the monitor satisfies $r_k^T \Pi r_k \leq 1 \; \forall \; k \geq k^*$ and r_k solution of (15).*

In the appendix, we give tools for obtaining a matrix Π satisfying Assumption 1 for a desired k^* as a function of the initial estimation error e_1 and a desired *tightness level* of the ellipsoidal bound.

3.4 Dynamic Output Feedback Controller

We consider general dynamic output feedback controllers of the form:

$$\begin{cases} x_{k+1}^c = A^c x_k^c + B^c \bar{y}_k, \\ u_k = C^c x_k^c + D^c \bar{y}_k, \end{cases} \tag{17}$$

with controller state $x^c \in \mathbb{R}^n$, networked output $\bar{y}$, control input u, and controller matrices (A^c, B^c, C^c, D^c) of appropriate dimensions. For simplicity, we only consider controllers with the same order as the plant. This is particulary important in the synthesis section of the manuscript (however, results for general order controllers can be derived following the same approach). The closed-loop system (11),(12),(17) can be written in terms of the estimation error $e_k = x_k - \hat{x}_k$ as follows:

$$\begin{cases} x_{k+1}^p = (A^p + B^p D^c C^p) x_k^p + B^p C^c x_k^c \\ \qquad\quad + B^p D^c F \eta_k + E v_k + B^p D^c \Gamma \delta_k, \\ x_{k+1}^c = A^c x_k^c + B^c C^p x_k^p + B^c F \eta_k + B^c \Gamma \delta_k, \\ e_{k+1} = (A^p - L C^p) e_k - L F \eta_k + E v_k - L \Gamma \delta_k. \end{cases} \tag{18}$$

4 Analysis Tools: Attacker's Reachable Sets

In this section, we provide tools for *quantifying* (for given (L, A^c, B^c, C^c, D^c)) and *minimizing* (by redesigning (L, A^c, B^c, C^c, D^c)) the impact of the attack δ_k on the state of the system when the monitor (16) is used for attack detection. We are interested in attacks that keep the monitor from raising alarms. This class of attacks is what we refer to as *stealthy attacks*. Here, we characterize *ellipsoidal bounds* on the set of states that stealthy attacks can induce in the system. In particular, we provide tools based on Linear Matrix Inequalities (LMIs) for computing ellipsoidal bounds on the *reachable set* of the attack sequence given the system dynamics, the control strategy, the system monitor, and the set of sensors being attacked.

Assumption 2 *We assume that the attack to system (11),(12),(17) starts at $k = k^*$ (the monitor convergence time), i.e., the system has been operating without attacks for sufficiently long time so that the residual trajectories r_k, for $k \geq k^*$, are contained in the monitor ellipsoid $\{r \in \mathbb{R}^m | r^T \Pi r \leq 1\}$ before an attack occurs.*

We remark again that, using the results in Appendix A, we can design monitor matrices Π satisfying Assumption 1 for any desired k^* as a function of the initial estimation error, e_1, and a desired tightness level of the ellipsoidal bound. Thus, Assumption 2 is not conservative in the sense that k^* can be selected arbitrarily small by properly designing the corresponding monitor matrix $\Pi(k^*)$.

The attacker can compromise up to s sensors, $s \in \{1, \ldots, m\}$, of the system. Consider the monitor (16) and write z_k in terms of the estimation error e_k and δ_k:

$$z_k = r_k^T \Pi r_k = \left\| \Pi^{\frac{1}{2}} (C^p e_k + F \eta_k + \Gamma \delta_k) \right\|^2, \tag{19}$$

where $\Pi^{\frac{1}{2}}$ is the symmetric square root matrix of Π and $\|\cdot\|$ denotes Euclidian norm. The set of feasible attack sequences that the attacker can launch while satisfying $z_k \leq 1$ (i.e., without raising alarms by the monitor) can be written as the constrained control problem on δ_k:

$$\left\{ \delta_k \in \mathbb{R}^m \;\middle|\; \begin{array}{l} \left\| \Pi^{\frac{1}{2}} (C^p e_k + F \eta_k + \Gamma \delta_k) \right\|^2 \leq 1, \\ \text{and Eq. (18)}, \forall \; k \geq k^*, \end{array} \right\}. \tag{20}$$

Define the extended state $\zeta_k := ((x_k^p)^T, (x_k^c)^T, e_k^T)^T$. Given ζ_k at the starting attack instant, ζ_{k^*}, and the disturbance and attack sequences $\eta(\cdot) := \{\eta_1, \eta_2, \ldots\}$, $v(\cdot) := \{v_1, v_2, \ldots\}$, and $\delta(\cdot) := \{\delta_1, \delta_2, \ldots\}$, denote by $\psi_\delta^\zeta(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), \delta(\cdot))$ the solution of (18) for all $k \geq k^*$. Let $\psi_\delta^x(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), \delta(\cdot))$ be the partition of $\psi_\delta^\zeta(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), \delta(\cdot))$ corresponding to the plant trajectories, i.e., the solution x_k^p of (18). We are interested in the state trajectories that the attacker can induce in the system restricted to satisfy (20). To this end, we introduce the notion of *stealthy reachable set*.

Definition 2 *Given the attack selection matrix Γ, the stealthy reachable set $\mathcal{R}_{\Gamma,k}^x$ at time $k \geq k^*$, from the starting extended state ζ_{k^*}, is defined as the set of states, $\psi_\delta^x(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), \delta(\cdot))$, reachable by system (18) through all possible disturbances and attack sequences satisfying $\eta_k^T \eta_k \leq \bar{\eta}$, $v_k^T v_k \leq \bar{v}$, and (20), i.e.,*

$$\mathcal{R}_{\Gamma,k}^x := \left\{ x^p \in \mathbb{R}^n \;\middle|\; \begin{array}{l} x^p = \psi_\delta^x(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), \delta(\cdot)), \\ \zeta_{k^*} \in \mathbb{R}^{3n}, \delta_k, \zeta_k \text{ satisfy (20)}, \\ v_k^T v_k \leq \bar{v}, \eta_k^T \eta_k \leq \bar{\eta}, \forall \; k \geq k^* \end{array} \right\}. \tag{21}$$

In this manuscript, we propose to use the "size" of the set $\mathcal{R}_{\Gamma,k}^x$, in terms of volume (Lebesgue measure), as a *security metric*. Stealthy reachable sets provide insight of the system potential performance degradation induced by stealthy attacks. The volume of $\mathcal{R}_{\Gamma,k}^x$ quantifies the size of the portion of the state space that opponents could reach by tampering with particular subsets of sensors (i.e., as a function of the sensor selection matrix Γ). So, for different subsets of sensors being attacked, we have reachable sets of different sizes; and thus, one could quantify the system sensitivity to attacks in particular sensors by comparing their corresponding volumes. How-

ever, in general, it is not tractable to compute $\mathcal{R}^x_{\Gamma,k}$ exactly. Instead, we look for an outer approximation $\mathcal{E}^x_{\Gamma,k}$ satisfying $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^x_{\Gamma,k}$ for all $k \geq k^*$. In particular, for some positive definite $\mathcal{P}^x_\Gamma \in \mathbb{R}^{n\times n}$ and nonnegative function α^x_k, we look for *outer ellipsoidal approximations* of the form $\mathcal{E}^x_{\Gamma,k} = \{x^p \in \mathbb{R}^n | (x^p)^T \mathcal{P}^x_\Gamma x^p \leq \alpha^x_k\}$ such that $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^x_{\Gamma,k}$. That is, the ellipsoid $(x^p)^T\mathcal{P}^x_\Gamma x^p = \alpha^x_k$ contains all the possible trajectories that stealthy attacks of the form (20) can induce in the system. Because for LTI systems $\mathcal{E}^x_{\Gamma,k}$ is a good approximation of $\mathcal{R}^x_{\Gamma,k}$, and because $\mathcal{E}^x_{\Gamma,k}$ can be computed efficiently using LMIs, we use the volume of $\mathcal{E}^x_{\Gamma,k}$ as an approximation of the proposed security metric. This approximation allows us to quantify the potential "damage" that sensor attacks can induce to the system in terms of the set of sensors being compromised (the attacker's sensor selection matrix Γ). In Figure 3, we depict a schematic representation of the proposed ideas.

4.1 Analysis Tools

In (15), the residual is given by $r_k = C^p e_k + \Gamma\delta_k + F\eta_k$. Because Γ has full column rank by construction, we can write the attack sequence as $\delta_k = \Gamma^+(r_k - C^p e_k - F\eta_k)$, where Γ^+ denotes the Moore-Penrose inverse of Γ, and the closed-loop dynamics (18) as

$$\begin{aligned} x^p_{k+1} &= (A^p + B^p D^c C^p)x^p_k + B^p C^c x^c_k - B^p D^c \Gamma\Gamma^+ C^p e_k \\ &\quad + B^p D^c (I_m - \Gamma\Gamma^+)F\eta_k + Ev_k + B^p D^c \Gamma\Gamma^+ r_k, \quad (22)\\ x^c_{k+1} &= A^c x^c_k + B^c C^p x^p_k - B^c \Gamma\Gamma^+ C^p e_k \\ &\quad + B^c (I_m - \Gamma\Gamma^+)F\eta_k + B^c \Gamma\Gamma^+ r_k, \quad (23)\\ e_{k+1} &= \big(A^p - L(I_m - \Gamma\Gamma^+)C^p\big)e_k \\ &\quad - L(I_m - \Gamma\Gamma^+)F\eta_k + Ev_k - L\Gamma\Gamma^+ r_k. \quad (24) \end{aligned}$$

Define the matrices:

$$\left\{ \begin{aligned} \mathcal{A} &:= \begin{bmatrix} A^p + B^p D^c C^p & B^p C^c & -B^p D^c \Gamma\Gamma^+ C^p \\ B^c C^p & A^c & -B^c \Gamma\Gamma^+ C^p \\ \mathbf{0} & \mathbf{0} & A^p - L(I_m - \Gamma\Gamma^+)C^p \end{bmatrix}, \\ \mathcal{B}^1 &:= \begin{bmatrix} B^p D^c (I_m - \Gamma\Gamma^+)F \\ B^c (I_m - \Gamma\Gamma^+)F \\ -L(I_m - \Gamma\Gamma^+)F \end{bmatrix}, \mathcal{B}^2 := \begin{bmatrix} E \\ \mathbf{0} \\ E \end{bmatrix}, \\ \mathcal{B}^3 &:= \begin{bmatrix} B^p D^c \Gamma\Gamma^+ \\ B^c \Gamma\Gamma^+ \\ -L\Gamma\Gamma^+ \end{bmatrix}, \mathcal{B} := \begin{bmatrix} \mathcal{B}^1 & \mathcal{B}^2 & \mathcal{B}^3 \end{bmatrix}. \end{aligned} \right. \quad (25)$$

Then, the closed-loop dynamics can be written in terms of the extended state $\zeta_k = ((x^p_k)^T, (x^c_k)^T, e^T_k)^T$:

$$\zeta_{k+1} = \mathcal{A}\zeta_k + \mathcal{B}^1\eta_k + \mathcal{B}^2 v_k + \mathcal{B}^3 r_k, \ k \geq k^*. \quad (26)$$

Denote by $\psi^\zeta_r(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), r(\cdot))$ the solution of (26) at time $k \geq k^*$ given the extended state at the starting attack instant ζ_{k^*} and the infinite *residual* and disturbance sequences $r(\cdot) := \{r_1, r_2, \ldots\}$, $\eta(\cdot)$, and $v(\cdot)$. Define the reachable set:

$$\mathcal{R}^\zeta_{\Gamma,k} := \left\{ \zeta \in \mathbb{R}^{3n} \left| \begin{array}{l} \zeta = \psi^\zeta_r(k, \zeta_{k^*}, \eta(\cdot), v(\cdot), r(\cdot)), \\ \zeta_{k^*} \in \mathbb{R}^{3n}, r^T_k \Pi r_k \leq 1, \\ v^T_k v_k \leq \bar{v}, \eta^T_k \eta_k \leq \bar{\eta}, \forall\, k \geq k^*. \end{array} \right. \right\}. \quad (27)$$

The set $\mathcal{R}^\zeta_{\Gamma,k}$ is the reachable set of an LTI system driven by peak-bounded perturbations. Therefore, we can use Corollary 1 to obtain outer approximations of the form $\mathcal{E}^\zeta_{\Gamma,k} = \{\zeta \in \mathbb{R}^{3n} | \zeta^T \mathcal{P}^\zeta_\Gamma \zeta \leq \alpha^\zeta_k\}$ such that $\mathcal{R}^\zeta_{\Gamma,k} \subseteq \mathcal{E}^\zeta_{\Gamma,k}$.

Remark 3 *We are ultimately interested in the stealthy reachable set of the plant states $\mathcal{R}^x_{\Gamma,k}$ introduced in (21). Note that $\mathcal{R}^x_{\Gamma,k}$ is the projection of $\mathcal{R}^\zeta_{\Gamma,k}$ onto the x^p-hyperplane. Hence, if $\mathcal{R}^\zeta_{\Gamma,k} \subseteq \mathcal{E}^\zeta_{\Gamma,k}$, then $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^\zeta_{\Gamma,k}||_{x^p} =: \mathcal{E}^x_{\Gamma,k}$, where $\mathcal{E}^\zeta_{\Gamma,k}||_{x^p}$ denotes the projection of $\mathcal{E}^\zeta_{\Gamma,k}$ onto the x^p-hyperplane. Therefore, to obtain the ellipsoid $\mathcal{E}^x_{\Gamma,k}$ containing $\mathcal{R}^x_{\Gamma,k}$, we can first obtain $\mathcal{E}^\zeta_{\Gamma,k}$ containing $\mathcal{R}^\zeta_{\Gamma,k}$ and then take $\mathcal{E}^\zeta_{\Gamma,k}||_{x^p}$ to obtain $\mathcal{E}^x_{\Gamma,k}$.*

Theorem 1 *Consider the closed-loop dynamics (22)-(24) with system matrices (A^p, B^p, C^p), observer gain L, controller matrices (A^c, B^c, C^c, D^c), monitor matrix Π, perturbations bounds $\bar{v}, \bar{\eta} \in \mathbb{R}_{>0}$, and attack selection matrix Γ. For a given $a \in (0,1)$, if there exist constants $a_1 = a^*_1, \ldots, a_N = a^*_N$ and matrix $\mathcal{P} = \mathcal{P}^*$ solution of (8) with $A = \mathcal{A}$, $N = 3$, $B^1 = \mathcal{B}^1$, $B^2 = \mathcal{B}^2$, $B^3 = \mathcal{B}^3$, $(\mathcal{A}, \mathcal{B})$ as defined in (25), $W_1 = (1/\bar{\eta})I_m$, $W_2 = (1/\bar{v})I_n$, $W_3 = \Pi$, $p_1 = m$, $p_2 = n$, and $p_3 = m$; then, for $k \geq k^*$, $\mathcal{R}^\zeta_{\Gamma,k} \subseteq \mathcal{E}^\zeta_{\Gamma,k} := \{\zeta \in \mathbb{R}^{3n} | \zeta^T \mathcal{P}^\zeta_\Gamma \zeta \leq \alpha^\zeta_k\}$, with $\mathcal{P}^\zeta_\Gamma := \mathcal{P}^*$ and $\alpha^\zeta_k := a^{k-1}\zeta^T_{k^*}\mathcal{P}^*\zeta_{k^*} + \big((3-a)(1-a^{k-1})\big)/(1-a)$, and the ellipsoid $\mathcal{E}^\zeta_{\Gamma,k}$ has minimum volume in the sense of Corollary 1.*

Proof: Consider the reachable set $\mathcal{R}^\zeta_{\Gamma,k}$ in (27). The set $\mathcal{R}^\zeta_{\Gamma,k}$ is the reachable set of system (26), which is a LTI system driven by peak-bounded perturbations. It follows that, under the conditions stated in Theorem 1, Corollary 1 can be used to obtain outer ellipsoidal approximations of the form $\mathcal{E}^\zeta_{\Gamma,k} = \{\zeta \in \mathbb{R}^{3n} | \zeta^T \mathcal{P}^\zeta_\Gamma \zeta \leq \alpha^\zeta_k\}$ such that $\mathcal{R}^\zeta_{\Gamma,k} \subseteq \mathcal{E}^\zeta_{\Gamma,k}$, where the sequence α^ζ_k is given by $\alpha^\zeta_k = a^{k-1}\zeta^T_{k^*}\mathcal{P}^*\zeta_{k^*} + (3-a)(1-a^{k-1})/(1-a)$, $\mathcal{P}^\zeta_\Gamma = \mathcal{P}^*$, and $\mathcal{P}^*$ is the solution of the optimization problem (8). The volume of $\mathcal{E}^\zeta_{\Gamma,k}$ is minimal in the sense of Corollary 1 because we solve (8) to obtain $\mathcal{P}^*$. ■

If the conditions of Theorem 1 are satisfied, for every $k \geq k^*$, the trajectories of the extended dynamics (26) are contained in $\mathcal{E}^\zeta_{\Gamma,k}$. Having this ellipsoid, we look for the projection $\mathcal{E}^\zeta_{\Gamma,k}||_{x^p}$ to obtain the ellipsoidal approximation $\mathcal{E}^x_{\Gamma,k} = \{x^p \in \mathbb{R}^n | (x^p)^T \mathcal{P}^x_\Gamma x^p \leq \alpha^x_k\}$ such that $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^x_{\Gamma,k}$. We use Lemma 10 in the Appendix to obtain this projection.

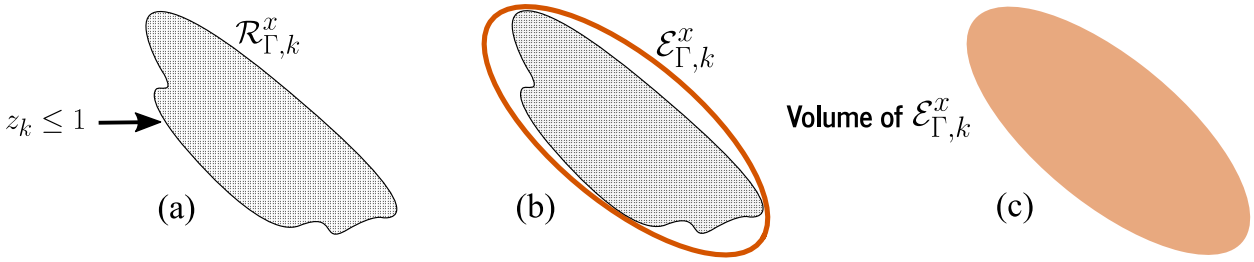

Fig. 3. (a) Stealthy reachable set $\mathcal{R}^x_{\Gamma,k}$; (b) ellipsoidal outer approximation $\mathcal{E}^x_{\Gamma,k}$ of $\mathcal{R}^x_{\Gamma,k}$; and (c) the volume of $\mathcal{E}^x_{\Gamma,k}$ as an approximation of the security metric (the volume of $\mathcal{R}^x_{\Gamma,k}$).

Corollary 2 *Let the conditions of Theorem 1 be satisfied and consider the corresponding matrix $\mathcal{P}^\zeta_\Gamma$ and function α^ζ_k. Let $\mathcal{P}^\zeta_\Gamma$ be partitioned as*

$$\mathcal{P}^\zeta_\Gamma =: \begin{bmatrix} \mathcal{P}^\zeta_1 & \mathcal{P}^\zeta_2 \\ (\mathcal{P}^\zeta_2)^T & \mathcal{P}^\zeta_3 \end{bmatrix},$$

with $\mathcal{P}^\zeta_1 \in \mathbb{R}^{n\times n}$, $\mathcal{P}^\zeta_2 \in \mathbb{R}^{n\times 2n}$, and $\mathcal{P}^\zeta_3 \in \mathbb{R}^{2n\times 2n}$. Then, for $k \geq k^*$, $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^x_{\Gamma,k} := \{x^p \in \mathbb{R}^n | (x^p)^T \mathcal{P}^x_\Gamma x^p \leq \alpha^x_k\}$ with $\mathcal{P}^x_\Gamma := \mathcal{P}^\zeta_1 - \mathcal{P}^\zeta_2 (\mathcal{P}^\zeta_3)^{-1} (\mathcal{P}^\zeta_2)^T$ and $\alpha^x_k := \alpha^\zeta_k$.

Proof: *By Theorem 1, the trajectories of (26) satisfy $\zeta_k^T \mathcal{P}^\zeta_\Gamma \zeta_k \leq \alpha^\zeta_k$ for $k \geq k^*$. By Lemma 10 in the Appendix, the projection of $\zeta_k^T \mathcal{P}^\zeta_\Gamma \zeta_k \leq \alpha^\zeta_k$ onto the x^p-hyperplane is given by $\mathcal{E}^x_{\Gamma,k}$ defined above. Thus, in light of Remark 3, the trajectories of the plant dynamics are contained in $\mathcal{E}^x_{\Gamma,k}$, i.e., $\mathcal{R}^x_{\Gamma,k} \subseteq \mathcal{E}^x_{\Gamma,k}$ for all $k \geq k^*$.* ■

4.2 Distance to Critical States: Analysis

The first security metric (the volume of $\mathcal{R}^x_{\Gamma,k}$) tell us the size of the part of the state space that opponents could access as a function of the sensors being attacked. So, we are implicitly assuming that all points in the state space are equally important (in terms of security) and we are just interested in the overall "number" of states potentially reachable by attacks. However, if there are regions of the state space that are more important than others (again in terms of security), and thus we are particularly interested in knowing if these regions are reachable by attacks, we need a different metric. To this end, as a second security metric, we propose to use the minimum distance between $\mathcal{R}^x_{\Gamma,k}$ and a possible set of *critical states* $\mathcal{C}^x$ – states that, if reached, compromise the integrity or safe operation of the system. Such a region might represent states in which, for example, the pressure of a holding vessel exceeds its pressure rating or the level of a liquid in a tank exceeds its capacity. However, because $\mathcal{R}^x_{\Gamma,k}$ is not known exactly, this distance cannot be directly computed. Instead, once the ellipsoidal bound $\mathcal{E}^x_{\Gamma,k}$ on $\mathcal{R}^x_{\Gamma,k}$ is obtained, we compute the minimum distance $d^x_{\Gamma,k}$ from $\mathcal{E}^x_{\Gamma,k}$ to $\mathcal{C}^x$ and use this $d^x_{\Gamma,k}$ as an approximation of the distance between $\mathcal{R}^x_{\Gamma,k}$ and $\mathcal{C}^x$ in terms of the set of sensors being compromised (the attacker's sensor selection matrix Γ). The distance $d^x_{\Gamma,k}$ gives us intuition of how far the actual reachable set $\mathcal{R}^x_{\Gamma,k}$ is from $\mathcal{C}^x$.

The set of critical states in many practical applications can be captured through the union of half-spaces defined by their boundary hyperplanes:

$$\mathcal{C}^x := \left\{ x^p \in \mathbb{R}^n \;\middle|\; \bigcup_{i=1}^{N} c_i^T x^p \geq b_i \right\}, \tag{28}$$

where each pair (c_i, b_i), $c_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, N$ quantifies a hyperplane that defines a single half-space.

Corollary 3 *Consider the set of critical states $\mathcal{C}^x$ defined in (28), and the matrix $\mathcal{P}^x_\Gamma$ and the function α^x_k obtained in Theorem 1. The minimum distance, $d^x_{\Gamma,k}$, between the outer ellipsoidal approximation of $\mathcal{R}^x_{\Gamma,k}$, $\mathcal{E}^x_{\Gamma,k} = \{x^p \in \mathbb{R}^n | (x^p)^T \mathcal{P}^x_\Gamma x^p \leq \alpha^x_k\}$, and $\mathcal{C}^x$ is given by*

$$d^x_{\Gamma,k} = \min \left(\frac{|b_i| - \sqrt{c_i^T (\mathcal{P}^x_\Gamma)^{-1} c_i / \alpha^x_k}}{c_i^T c_i} \right), \; i = 1, \ldots, N. \tag{29}$$

Proof: The minimum distance between an ellipsoid centered at the origin $\{x \in \mathbb{R}^n | x^T \mathcal{P} x = 1\}$, $\mathcal{P} \in \mathbb{R}^{n\times n}$, $\mathcal{P} > 0$ and a hyperplane $\{x \in \mathbb{R}^n | c^T x = b\}$, $c \in \mathbb{R}^n$, $b \in \mathbb{R}$ is given by the formula $(|b| - \sqrt{c^T \mathcal{P}^{-1} c})/c^T c$, [16,17]. It follows that the minimum distance between $\mathcal{D}^x$, conformed by the N hyperplanes in (28), and $\mathcal{E}^x_{\Gamma,k}$ is simply given by $d^x_{\Gamma,k}$ in (29). ■

Remark 4 *If $d^x_{\Gamma,k} > 0$, the ellipsoid $\mathcal{E}^x_{\Gamma,k}$ bounding $\mathcal{R}^x_{\Gamma,k}$ and the set of critical states $\mathcal{C}^x$ do not intersect; if $d^x_{\Gamma,k} = 0$, they touch at a point only; and $d^x_{\Gamma,k} < 0$ implies that they intersect. In Figure 4, we depict a schematic representation of these ideas. Note that, due to potential conservatism of the ellipsoidal bounds, $d^x_{\Gamma,k} < 0$ does not necessarily imply that $\mathcal{R}^x_{\Gamma,k}$ and $\mathcal{C}^x$ intersect (see Figure 4 (d)). However, $d^x_{\Gamma,k} \geq 0$ does imply that they do not intersect, which is advantageous from the security perspective. Then, if we secure sensors leading to $d^x_{\Gamma,k} < 0$ or redesign controllers and monitors such that $d^x_{\Gamma,k} \geq 0$, we ensure that $\mathcal{R}^x_{\Gamma,k}$ and $\mathcal{C}^x$ does not intersect.*

Remark 5 *Note that it might be possible that perturbations alone drive the system to critical states even without attacks. However, it is the effect of both together, attacks and perturbations, that we want to quantify. For*

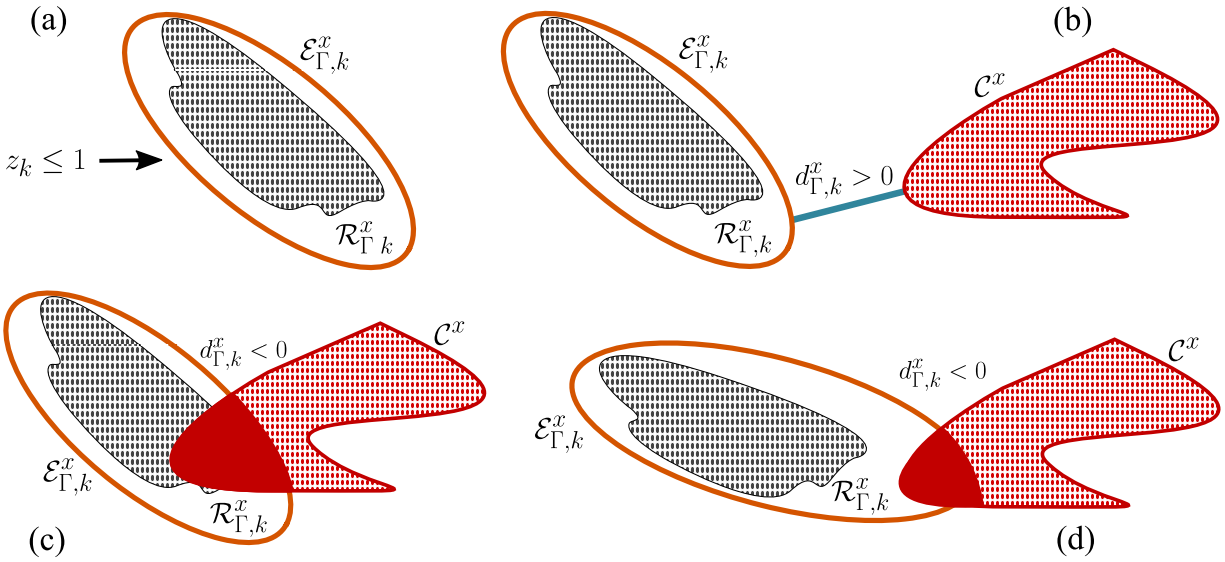

Fig. 4. (a) Stealthy reachable set $\mathcal{R}^x_{\Gamma,k}$ and ellipsoidal outer approximation $\mathcal{E}^x_{\Gamma,k}$ of $\mathcal{R}^x_{\Gamma,k}$; and (b)-(d) minimum distance $d^x_{\Gamma,k}$ between $\mathcal{E}^x_{\Gamma,k}$ and critical states $\mathcal{C}^x$.

instance, a possible scenario is that critical states are not reachable by perturbations alone, but when you combine attacks and perturbations, critical states might be reachable by some realizations of attacks and perturbations combined. Another scenario is that critical states are reachable by perturbations alone in the first place. In this case, if one adds attacks to the mixture (due to superposition of linear systems), there would exist realizations of attacks and perturbations that lead to a larger portion of critical states being reachable. It is this combined uncertainty of having attacks and unknown perturbations that we aim at quantifying in terms of security.

4.3 Simulation Results

Consider the closed-loop system (18) with matrices as in (30), $\bar{\eta} = \sqrt{\pi}$, and $\bar{v} = 1$. The controller matrices (A^c, B^c, C^c, D^c) are designed to guarantee that the $\mathcal{L}_2$-gain [37] from the vector of perturbations $(v_k^T, \eta_k^T)^T$ to the performance output $s_k = 0.25x_k^{p,3} + \eta_k^3$ is upper bounded by $\gamma = 3$. We use the results in the appendix to design the monitor matrix Π so that, for $k > k^* = 10$, $r_k \Pi r_k \leq 1$. Using Theorem 1, we obtain $\mathcal{E}^\zeta_{\Gamma,k}$ for all the possible combinations of the sensor attack selection matrix Γ. Once we have $\mathcal{E}^\zeta_{\Gamma,k}$, using Corollary 2, we project $\mathcal{E}^\zeta_{\Gamma,k}$ onto the x^p-hyperplane to obtain $\mathcal{E}^x_{\Gamma,k}$. Note that we have k-dependent approximations $\mathcal{E}^x_{\Gamma,k}$ of $\mathcal{R}^x_{\Gamma,k}$; however, because $a < 1$, the function α^x_k conforming $\mathcal{E}^x_{\Gamma,k}$ converge exponentially to $(3-a)/(1-a)$. It follows that, in a few time steps, $\mathcal{E}^x_{\Gamma,k} \approx \mathcal{E}^x_{\Gamma,\infty} = \{x \in \mathbb{R}^n | x^T \mathcal{P}^x_\Gamma x \leq (3-a)/(1-a)\}$, and thus, $\mathcal{E}^x_{\Gamma,k} \approx \mathcal{E}^x_{\Gamma,\infty}$. We present $\mathcal{E}^x_{\Gamma,\infty}$ instead of the time-dependent $\mathcal{E}^x_{\Gamma,k}$. In Figure 5, we show the projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,2}, x^{p,3})$-hyperplane for different sets of sensor being attacked. Figure 6 depicts the projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,1}, x^{p,2})$-hyperplane and the distance to the set of critical states $\mathcal{C}^x = \{x^p \in \mathbb{R}^3 | x^{p,1} \leq -15\}$. In Table 1, we give the numerical values of the volume of $\mathcal{E}^x_{\Gamma,\infty}$ and the distance to the critical states depicted in Figure 6 for different sensors being attacked. Note that some distances are negative, as explained in Remark 4, negative distances imply that there is a nonempty intersection between the critical states and the stealthy reachable set. That is, there exist attack sequences that can drive the system to the unsafe region without being detected by the system monitor. Assume, for instance, that two out of the three sensors can be completely secured, i.e., attacks to these sensors are impossible. From Table 1, we note that attacks to sensor two leads to the largest volume of $\mathcal{E}^x_{\Gamma,\infty}$ and the smallest distance to critical states $d^x_{\Gamma,\infty}$. Therefore, if only two sensors can be secured, they should be sensors two and three. Following the same logic, if only one sensor can be secured, then sensor two must be selected because attacks to the remaining sensors, one and three, lead to the smallest $\mathcal{E}^x_{\Gamma,\infty}$ and the largest $d^x_{\Gamma,\infty}$. Thereby, our tools can be used to allocate security equipment to sensors when limited resources are available.

5 Synthesis Tools: Attacker's Reachable Sets

Next, we derive tools for designing the monitor and controller matrices $\kappa := (L, \Pi, A^c, B^c, C^c, D^c)$ such that the impact of stealthy attacks on the system dynamics is minimized. We first design κ to minimize the volume of $\mathcal{E}^x_{\Gamma,k}$ (thus decreasing the size of $\mathcal{R}^x_{\Gamma,k}$) while guaranteeing some attack-free prescribed performance of the closed-loop system.

Attacked Sensors	Volume of $\mathcal{E}^x_{\Gamma,\infty}$	Distance to Critical States $d^x_{\Gamma,\infty}$
(1)	150.72	8.07
(2)	453.51	4.20
(3)	219.43	8.60
(1,2)	952.95	-2.38
(1,3)	279.50	6.85
(2,3)	2063.46	-6.67
(1,2,3)	4300.32	-23.01

Table 1
Volume of the approximation $\mathcal{E}^x_{\Gamma,\infty}$ of $\mathcal{R}^x_{\Gamma,\infty}$ and distance $d^x_{\Gamma,\infty}$ to the critical states $\mathcal{C}^x$ for different attacked sensors.

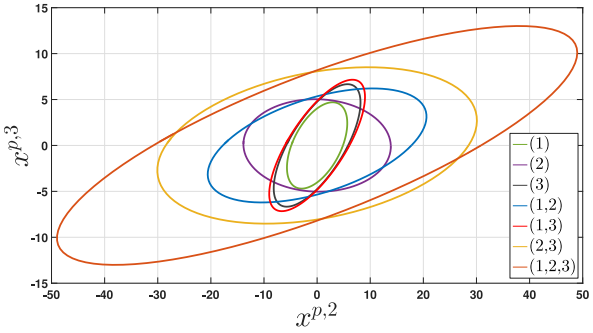

Fig. 5. Projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,2}, x^{p,3})$-hyperplane for different sets of sensor being attacked.

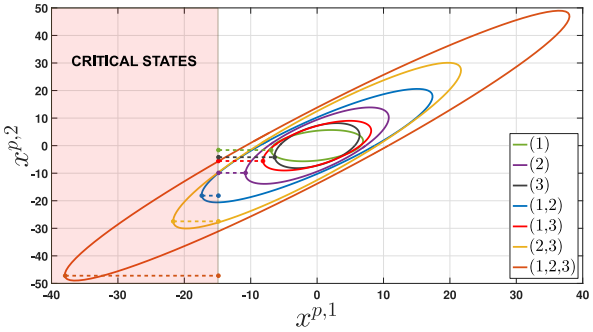

Fig. 6. Projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,1}, x^{p,2})$-hyperplane for different sets of sensor being attacked and distance to critical states.

Remark 6 *We present synthesis results in terms of the sensor attack selection matrix Γ. That is, for given Γ, we provide synthesis tools to design optimal controllers and monitors – optimal in terms of minimal volume $\mathcal{E}^x_{\Gamma,\infty}$ for a desired attack-free closed-loop system performance. Note, however, that we do not have access to Γ in practice, i.e., because we assume stealthy attacks, the set of sensors being attacked is unknown to the system designer. Nevertheless, once we have derived synthesis results for given Γ, we provide general guidelines for using these results to synthesize controllers/monitors for unknown matrix Γ. In particular, we propose techniques from sensor protection placement in power systems* [9,15]*; and game-theoretic techniques* [3].

Consider the extended attacker's reachable set $\mathcal{R}^\zeta_{\Gamma,k}$ defined in (27) with matrices $(\mathcal{A}, \mathcal{B})$ as in (25). Note that, for every realization of $\kappa = (L, \Pi, A^c, B^c, C^c, D^c)$, using Theorem 1 and Corollary 2, we can obtain $\mathcal{E}^x_{\Gamma,k}$ containing $\mathcal{R}^x_{\Gamma,k}$. Here, we aim at finding the $\kappa = \kappa^*$ leading to the smallest possible volume of $\mathcal{E}^x_{\Gamma,\infty}$ (see (7)) among all realizations of $(L, \Pi, A^c, B^c, C^c, D^c)$. If we let κ be optimization variables rather than given parameters, by Proposition 1, to find κ^*, we have to find (L, A^c, B^c, C^c, D^c) conforming the matrices $(\mathcal{A}, \mathcal{B})$, the constants (a_1, a_2, b), and the matrices $\mathcal{P}$ and Π solution of the optimization problem:

$$\begin{cases} \min\limits_{\kappa,\mathcal{P},a_1,a_2,b} \; -\log\det[\mathcal{P}], \\ \text{s.t. } a_1, a_2, b \in (0,1),\ a_1 + a_2 + b \geq a,\ \mathcal{P} > 0, \text{ and} \\ \mathcal{L} := \begin{bmatrix} a\mathcal{P} & \mathcal{A}^T\mathcal{P} & \mathbf{0} \\ \mathcal{P}\mathcal{A} & \mathcal{P} & \mathcal{P}\mathcal{B} \\ \mathbf{0} & \mathcal{B}^T\mathcal{P} & W_{a_i} \end{bmatrix} \geq \mathbf{0}; \end{cases} \tag{31}$$

with $W_{a_i} := \text{diag}[\frac{1-a_1}{\bar{\eta}} I_m, \frac{1-a_2}{\bar{v}} I_n, (1-b)\Pi]$. However, because $(L, \Pi, A^c, B^c, C^c, D^c)$ are now variables, the blocks $\mathcal{P}\mathcal{A}$, $\mathcal{P}\mathcal{B}$, and $b\Pi$ in (31) are nonlinear in $(\kappa, \mathcal{P})$. Following the results in [31], we propose an invertible linearizing change of variables:

$$(\mathcal{P}, \kappa) \to \nu := ((X, Y, S), (R, G), (K, O, M, N)), \tag{32}$$

such that, in the new variables ν, the objective in (31) is convex and the restrictions are affine. In particular, for $\mathcal{P} > 0$ and the *nonlinear* matrix inequality $\mathcal{L} \geq 0$ defined in (31), we aim at finding two invertible matrices $\mathcal{T}_1$ and $\mathcal{T}_2$ such that the congruence transformations $\mathcal{P} \to \mathcal{T}_1^T \mathcal{P} \mathcal{T}_1$ and $\mathcal{L} \to \mathcal{T}_2^T \mathcal{L} \mathcal{T}_2$ lead to new *linear* matrix inequalities $\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 > \mathbf{0}$ and $\mathcal{T}_2^T \mathcal{L} \mathcal{T}_2 \geq \mathbf{0}$ in ν.

5.1 Change of Variables and Optimization Problem

To be able to convexify the synthesis problem, we have to impose some block diagonal structure on the matrix $\mathcal{P}$. Let $\mathcal{P}$ be positive definite and of the form

$$\mathcal{P} := \begin{bmatrix} X & U & \mathbf{0} \\ U^T & \tilde{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S \end{bmatrix}, \tag{33}$$

with $X, U, \tilde{X}, S \in \mathbb{R}^{n \times n}$ and positive definite X, $\tilde{X}$, and S. Define the matrices:

$$\mathcal{X} := \begin{bmatrix} X & U \\ U^T & \tilde{X} \end{bmatrix}, \mathcal{X}^{-1} =: \begin{bmatrix} Y & V \\ V^T & \tilde{Y} \end{bmatrix}, \mathcal{Y} := \begin{bmatrix} Y & I \\ V^T & \mathbf{0} \end{bmatrix}, \mathcal{Z} := \begin{bmatrix} I & \mathbf{0} \\ XU \end{bmatrix}. \tag{34}$$

Using block matrix inversion formulas, it is easy to verify that $YX + VU^T = I$ and $YU + V\tilde{X} = \mathbf{0}$, which leads to $\mathcal{Y}^T \mathcal{X} = \mathcal{Z}$. Define the matrices $\mathcal{T}_1$ and $\mathcal{T}_2$ as

$$\mathcal{T}_1 := \begin{bmatrix} \mathcal{Y} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} = \begin{bmatrix} Y & I & \mathbf{0} \\ V^T & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \end{bmatrix} \in \mathbb{R}^{3n \times 3n}, \tag{35}$$

$$\mathcal{T}_2 := \begin{bmatrix} \mathcal{T}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I \end{bmatrix} \in \mathbb{R}^{9n \times 9n}. \tag{36}$$

$$
\begin{cases}
\left(\begin{array}{c|c} A^p & B^p \\ \hline C^p & D^p \end{array}\right) = \left(\begin{array}{ccc|cc} 0.62 & 0.21 & 0.03 & 0.07 & 1.0 \\ 0.08 & 0.72 & 0.54 & 0.23 & 0.5 \\ 0.02 & 0.02 & 0.65 & 0 & 1.0 \\ \hline 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right), \left(\begin{array}{c|c} A^c & B^c \\ \hline C^c & D^c \end{array}\right) = \left(\begin{array}{ccc|ccc} 0.10 & 0.09 & -0.16 & -0.24 & 0.10 & 0.24 \\ -0.06 & -0.06 & 0.09 & 0.06 & -0.06 & -0.06 \\ -0.08 & -0.07 & 0.08 & 0.12 & -0.07 & -0.15 \\ \hline -0.08 & 1.38 & 0.85 & -0.51 & -1.74 & 0.01 \\ 0.09 & -0.08 & 0.12 & -0.14 & -0.09 & -0.27 \end{array}\right), \\
L = \begin{pmatrix} 0.52 & 0.21 & 0.03 \\ 0.08 & 0.52 & 0.54 \\ 0.02 & 0.02 & 0.35 \end{pmatrix}, \Pi = \begin{pmatrix} 9.50 & -0.76 & -0.05 \\ -0.76 & 7.69 & -0.95 \\ -0.05 & -0.95 & 8.14 \end{pmatrix} \times 10^{-2}, E = I_n, F = I_m.
\end{cases} \tag{30}
$$

Then, $\mathcal{P} \to \mathcal{T}_1^T \mathcal{P} \mathcal{T}_1$ and $\mathcal{L} \to \mathcal{T}_2^T \mathcal{L} \mathcal{T}_2$ take the form:

$$
\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 = \begin{bmatrix} Y & I & \mathbf{0} \\ I & X & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & S \end{bmatrix} =: \mathbf{P}(\nu), \tag{37}
$$

$$
\mathcal{T}_2^T \mathcal{L} \mathcal{T}_2 = \begin{bmatrix} a\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 & \mathcal{T}_1^T \mathcal{A}^T \mathcal{P} \mathcal{T}_1 & \mathbf{0} \\ \mathcal{T}_1^T \mathcal{P} \mathcal{A} \mathcal{T}_1 & \mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 & \mathcal{T}_1^T \mathcal{P} \mathcal{B} \\ \mathbf{0} & \mathcal{B}^T \mathcal{P} \mathcal{T}_1 & W_{a_i} \end{bmatrix}. \tag{38}
$$

The structure of $\mathbf{P}(\nu)$ follows from symmetry of $\mathcal{P}$, which implies symmetric X and Y and $XY + UV^T = I$. Note that the block $\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1$ is linear in X, Y, and S. Next, using the definition of $(\mathcal{A}, \mathcal{B})$ in (25), we expand the blocks $\mathcal{T}_1^T \mathcal{P} \mathcal{A} \mathcal{T}_1$ and $\mathcal{T}_1^T \mathcal{P} \mathcal{B}$. Note that the matrix $\mathcal{A}$ is upper triangular. Let $\mathcal{A}$ be partitioned as

$$
\mathcal{A} =: \begin{bmatrix} \mathcal{A}_1 & \mathcal{A}_2 \\ \mathbf{0} & \mathcal{A}_3 \end{bmatrix}; \tag{39}
$$

and define the change of controller, observer, and monitor variables:

$$
\begin{pmatrix} K - XA^pY & O \\ M & N \end{pmatrix} := \begin{pmatrix} U & XB^p \\ \mathbf{0} & I_l \end{pmatrix} \begin{pmatrix} A^c & B^c \\ C^c & D^c \end{pmatrix} \times \begin{pmatrix} V^T & \mathbf{0} \\ C^pY & I_m \end{pmatrix}, \tag{40a}
$$

$$
R := SL, \tag{40b}
$$

$$
G := \Pi. \tag{40c}
$$

Then, $\mathcal{T}_1^T \mathcal{P} \mathcal{A} \mathcal{T}_1$ can be written as

$$
\mathbf{A}(\nu) := \mathcal{T}_1^T \mathcal{P} \mathcal{A} \mathcal{T}_1 = \begin{bmatrix} \mathcal{Y}^T \mathcal{X} \mathcal{A}_1 \mathcal{Y} & \mathcal{Z} \mathcal{A}_2 \\ \mathbf{0} & S\mathcal{A}_3 \end{bmatrix} \tag{41}
$$

$$
= \begin{bmatrix} A^pY + B^pM & A^p + B^pNC^p & -B^pN\Gamma\Gamma^+C^p \\ K & XA^p + OC^p & -O\Gamma\Gamma^+C^p \\ \mathbf{0} & \mathbf{0} & SA^p - R(I_m - \Gamma\Gamma^+)C^p \end{bmatrix},
$$

the block $\mathcal{T}_1^T \mathcal{P} \mathcal{B}$ as

$$
\mathbf{B}(\nu) := \mathcal{T}_1^T \mathcal{P} \mathcal{B} = \begin{bmatrix} \mathcal{Z} & \mathbf{0} \\ \mathbf{0} & S \end{bmatrix} \mathcal{B} \tag{42}
$$

$$
= \begin{bmatrix} B^pN(I_m - \Gamma\Gamma^+)F & E & B^pN\Gamma\Gamma^+ \\ O(I_m - \Gamma\Gamma^+)F & XE & O\Gamma\Gamma^+ \\ -R(I_m - \Gamma\Gamma^+)F & SE & -R\Gamma\Gamma^+ \end{bmatrix},
$$

and the block W_{a_i} as

$$
W_{a_i} = \operatorname{diag}\left[\frac{1 - a_1}{\bar{\eta}} I_m, \frac{1 - a_2}{\bar{v}} I_n, (1 - b)G \right] =: \mathbf{W}(\nu). \tag{43}
$$

Therefore, under $\mathcal{T}_1$, $\mathcal{T}_2$, and the new variables in (40), the blocks transforms as

$$
\begin{cases} \mathcal{P} \to \mathbf{P}(\nu), \;\; \mathcal{T}_1^T \mathcal{P} \mathcal{A} \mathcal{T}_1 \to \mathbf{A}(\nu), \\ \mathcal{T}_1^T \mathcal{P} \mathcal{B} \to \mathbf{B}(\nu), \;\; W_{a_i} \to \mathbf{W}(\nu), \end{cases} \tag{44}
$$

with $\mathbf{P}(\nu)$, $\mathbf{A}(\nu)$, $\mathbf{B}(\nu)$, and $\mathbf{W}(\nu)$ as defined in (37), (41), (42), and (43), respectively. That is, the original blocks, $\mathcal{P}\mathcal{A}$ and $\mathcal{P}\mathcal{B}$, that depend non-linearly on the decision variables $(\kappa, \mathcal{P})$ are transformed into blocks that are affine functions of the new variables ν. If ν is given and U and V are invertible, the change of variables in (40) and the matrix $\mathcal{T}_1$ are invertible and thus $(\kappa, \mathcal{P})$ can be constructed from ν and they are unique. Moreover, invertible V implies that $\mathcal{T}_1$ and $\mathcal{T}_2$ are nonsingular and thus the transformations $\mathcal{P} \to \mathcal{T}_1^T \mathcal{P} \mathcal{T}_1$ and $\mathcal{L} \to \mathcal{T}_2^T \mathcal{L} \mathcal{T}_2$ are congruent. The latter implies that

$$
\mathcal{P} > \mathbf{0} \text{ and } \mathcal{L} \geq \mathbf{0} \Leftrightarrow \mathbf{P}(\nu) > \mathbf{0} \text{ and } \mathbf{L}(\nu) \geq \mathbf{0}, \tag{45}
$$

where

$$
\mathbf{L}(\nu) := \mathcal{T}_2^T \mathcal{L} \mathcal{T}_2 = \begin{bmatrix} a\mathbf{P}(\nu) & \mathbf{A}(\nu)^T & \mathbf{0} \\ \mathbf{A}(\nu) & \mathbf{P}(\nu) & \mathbf{B}(\nu) \\ \mathbf{0} & \mathbf{B}(\nu)^T & \mathbf{W}(\nu) \end{bmatrix}. \tag{46}
$$

If the matrix $\mathbf{P}(\nu)$ is positive definite, by the Schur complement, $Y > 0$ and $X - Y^{-1} > 0$, and because $YX + VU^T = I$ by construction (see Eq. (34)), $VU^T = I - YX < \mathbf{0}$, i.e., the matrix VU^T is nonsingular. Therefore, if $\mathbf{P}(\nu) > 0$, it is always possible to find nonsingular U and V satisfying $YX + VU^T = I$. In the following lemma, we summarize the discussion presented above.

Lemma 2 *Consider the observer, monitor, and controller matrices $\kappa = (L, \Pi, A^c, B^c, C^c, D^c)$, and the matrices $\mathcal{L}$ and $\mathcal{P}$ as defined in (31) and (33), respectively. If there exists $\nu = (X, Y, S, R, G, K, O, M, N)$ satisfying $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$ with $\mathbf{P}(\nu)$ and $\mathbf{L}(\nu)$ as defined in (37) and (46), respectively; then, there exists $(\kappa, \mathcal{P})$ satisfying $\mathcal{P} > 0$ and $\mathcal{L} \geq \mathbf{0}$. Moreover, for every ν such that $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$, the change of variables in (40) and matrix $\mathcal{T}_1$ are invertible and the $(\kappa, \mathcal{P})$ obtained by inverting (37) and (40) is unique.*

Proof: *Assume that* ν *is such that* $\mathbf{P}(\nu) > \mathbf{0}$ *and* $\mathbf{L}(\nu) \geq \mathbf{0}$. *Because* $\mathbf{P}(\nu) > \mathbf{0}$, *by the Schur complement,* $Y > 0$ *and* $X - Y^{-1} > 0$. *Since* $YX + VU^T = I$, *then* $VU^T = I - YX < \mathbf{0}$, *i.e., the matrix* VU^T *is invertible. Hence, it is always possible to factorize* $I - YX$ *as* $VU^T = I - YX$ *with square and nonsingular* U *and* V. *Invertible* U *and* V *implies that* $\mathcal{T}_1$ *and* $\mathcal{T}_2$ *are square and nonsingular and thus the transformations* $\mathcal{P} \to \mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 = \mathbf{P}(\nu)$ *and* $\mathcal{L} \to \mathcal{T}_2^T \mathcal{L} \mathcal{T}_2 = \mathbf{L}(\nu)$ *are congruent. It follows that* $\mathbf{P}(\nu) > \mathbf{0}$ *and* $\mathbf{L}(\nu) \geq \mathbf{0}$ *imply* $\mathcal{P} > 0$ *and* $\mathcal{L} \geq \mathbf{0}$ *because* $\mathbf{P}(\nu)$ *and* $\mathbf{L}(\nu)$ *have the same signature as* $\mathcal{P}$ *and* $\mathcal{L}$, *respectively. Because* $\mathbf{P}(\nu) > \mathbf{0}$, *the matrices* U, V, *and* S *are nonsingular. This implies that the change of variables in* (40) *and* $\mathcal{T}_1$ *are invertible and lead to unique* $(\kappa, \mathcal{P})$ *by inverting* (37) *and* (40). ■

So far, we have derived from the analysis inequalities, $\mathcal{P} > \mathbf{0}$ and $\mathcal{L} \geq \mathbf{0}$ in (31), the synthesis inequalities $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$ defined in (37) and (46). If we find a realization of ν satisfying the synthesis inequalities, we factorize $I - YX$ into nonsingular matrices V and U satisfying $I - YX = VU^T$, use these V and U to solve the equations in (40) to obtain the controller, observer, and monitor matrices, and invert (37) to obtain the ellipsoid matrix $\mathcal{P}$. By Lemma 2, this $(\kappa, \mathcal{P})$ satisfies the analysis inequalities in (31).

We aim at minimizing the number of states that the attacker can induce in the system while remaining stealthy, i.e., we want to make the "size" of $\mathcal{R}^x_{\Gamma,k}$ defined in (21) as small as possible by selecting ν. To achieve this, we seek for the ν that minimizes the volume of $\mathcal{E}^x_{\Gamma,\infty}$ (which would decrease the size of $\mathcal{R}^x_{\Gamma,k}$). In the analysis case, we look for the matrix $\mathcal{P}$ satisfying $\mathcal{P} > \mathbf{0}$ and $\mathcal{L} \geq \mathbf{0}$ leading to the ellipsoid $\mathcal{E}^\zeta_{\Gamma,k} = \{\zeta \in \mathbb{R}^{3n} | \zeta^T \mathcal{P} \zeta \leq \alpha^\zeta_k\}$ bounding $\mathcal{R}^\zeta_{\Gamma,k}$ (defined in (27)) and then, using Corollary 2, we project this $\mathcal{E}^\zeta_{\Gamma,k}$ onto the x^p-hyperplane to obtain $\mathcal{E}^x_{\Gamma,k}$. To follow the same approach for synthesis, we would need to minimize the volume of $\zeta^T \mathcal{P} \zeta = \alpha^\zeta_\infty$ subject to $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$. However, the matrix $\mathcal{P}$ cannot be written in terms of ν and minimizing the volume of $\zeta^T \mathbf{P}(\nu) \zeta = \alpha^\zeta_\infty$ is not an equivalent objective. Instead, because the projection $\mathcal{E}^x_{\Gamma,k}$ can be written in terms of ν, we seek to minimize the volume of $\mathcal{E}^x_{\Gamma,\infty}$ directly.

Lemma 3 *Consider* $\mathcal{E}^\zeta_{\Gamma,k} = \{\zeta \in \mathbb{R}^{3n} | \zeta^T \mathcal{P} \zeta = \alpha^\zeta_k\}$ *with matrix* $\mathcal{P} \in \mathbb{R}^{3n \times 3n}$ *as defined in* (33), *extended state* $\zeta = ((x^p)^T, (x^c)^T, e^T)^T$, *and* $\alpha^\zeta_k \in \mathbb{R}_{>0}$, $k \in \mathbb{N}$. *The projection of* $\mathcal{E}^\zeta_{\Gamma,k}$ *onto the* x^p*-hyperplane is given by the ellipsoid* $\mathcal{E}^x_{\Gamma,k} = \{x^p \in \mathbb{R}^n | (x^p)^T Y^{-1} x^p = \alpha^\zeta_k\}$ *with* Y *as defined in* (34).

Proof: *For* $\mathcal{P}$ *as defined in* (33), *by Lemma 10 in the appendix, the boundary of the projection of* $\mathcal{E}^\zeta_{\Gamma,k}$ *onto the* x^p*-hyperplane,* $\mathcal{E}^x_{\Gamma,k}$, *is given by* $(x^p)^T (X - U\tilde{X}^{-1} U^T) x^p = \alpha^\zeta_k$. *Using standard block matrix inversion formulas (see, e.g., [13]) and the definition of* Y *in* (34), *we have* $Y = (X - U\tilde{X}^{-1}U^T)^{-1}$ *and therefore* $\mathcal{E}^x_{\Gamma,k}$ *can be written in terms of* ν *as* $\mathcal{E}^x_{\Gamma,k} = \{x^p \in \mathbb{R}^n | (x^p)^T Y^{-1} x^p = \alpha^\zeta_k\}$. ■

Lemma 3 implies that, in the new variables, we can minimize the volume of $(x^p)^T Y^{-1} x^p = \alpha^\zeta_\infty$ to reduce the size of $\mathcal{R}^x_{\Gamma,k}$. Therefore, in the synthesis case, we seek to minimize the volume of $(x^p)^T Y^{-1} x^p = \alpha^\zeta_\infty$ subject to $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$. The volume of $\mathcal{E}^x_{\Gamma,\infty}$ is proportional to $\sqrt{\det[Y]}$ for any $\alpha^\zeta_\infty > 0$ [16]. Moreover, the function $\sqrt{\det[Y]}$ shares the same minimizer with $\log\det[Y]$ [4]. However, the function $\log\det[Y]$ is concave for any positive definite matrix Y. To overcome this obstacle, we look for a convex upper bound on $\sqrt{\det[Y]}$ and minimize this bound instead. In order to derive this bound, we use the *Arithmetic Mean-Geometric Mean* (AM-GM) *Inequality* which states the following: For any sequence of positive real numbers, $c_1, c_2, \ldots, c_n$, the inequality $(\prod_{j=1}^n c_j)^{1/n} \leq \frac{1}{n} \sum_{j=1}^n c_j$ is satisfied [33].

Lemma 4 *For any positive definite matrix* $Y \in \mathbb{R}^{n \times n}$, *the following is satisfied:*

$$\det[Y]^{\frac{1}{n}} \leq \frac{1}{n} \mathrm{trace}[Y] \Rightarrow \det[Y]^{\frac{1}{2}} \leq \frac{1}{n^{\frac{n}{2}}} \mathrm{trace}[Y]^{\frac{n}{2}}. \tag{47}$$

Moreover, because Y *is positive definite*

$$\arg\min[\mathrm{trace}[Y]^{n/2}] = \arg\min[\mathrm{trace}[Y]];$$

that is, $\mathrm{trace}[Y]^{n/2}$ *and* $\mathrm{trace}[Y]$ *share the same minimizer. Therefore, by minimizing* $\mathrm{trace}[Y]$, *we minimize an upper bound on* $\sqrt{\det[Y]}$.

Proof: Let $\lambda_j[Y]$ denote the j-th eigenvalue of Y, $j = 1, \ldots, n$. Because Y is positive definite, the eigenvalues of Y are strictly positive. Then, because $\det[Y] = \prod_{j=1}^n \lambda_j[Y]$ and $\mathrm{trace}[Y] = \sum_{j=1}^n \lambda_j[Y]$, we have $(\prod_{j=1}^n \lambda_j[Y])^{1/n} \leq \frac{1}{n} \sum_{j=1}^n \lambda_j[Y]$ as a direct consequence of the (AM-GM) inequality [33], i.e., the left-hand side of (47) is satisfied for any positive definite Y. Given that both $\det[Y]$ and $\mathrm{trace}[Y]$ are strictly positive, the right-hand side of (47) follows from the left-hand side inequality by raising it to the power $n/2$. The function $g(x) := x^{n/2}$ is strictly positive and convex for $x > 0$. Hence, the upper bound $(1/n^{n/2})\mathrm{trace}[Y]^{n/2}$ in (47) is monotonically increasing in $\mathrm{trace}[Y]$. It follows that, for $Y > 0$, $\arg\min[\mathrm{trace}[Y]^{n/2}] = \arg\min[\mathrm{trace}[Y]]$ for any $n \in \mathbb{N}$, and the assertion follows. ■

Up to this point, we have the necessary tools for selecting ν to reduce the size of the stealthy reachable set $\mathcal{R}^x_{\Gamma,k}$. That is, we have the constraints, $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$, and the cost function, $\mathrm{trace}[Y]$, needed to cast the optimization problem to minimize the volume of $\mathcal{E}^x_{\Gamma,\infty}$. There is, however, one last ingredient to be considered before casting the complete synthesis optimization problem; namely, the attack-free performance of the closed-loop dynamics.

5.2 Attack-Free Observer, Monitor, and Controller Performance

As we now move towards posing the complete syntheses optimization problem, we note that as $||L|| \to 0$, $||B^c|| \to 0$, and $||D^c|| \to 0$, the reachable set $\mathcal{R}^x_{\Gamma,k}$ converges to the empty set because the attack-dependent terms in (18) vanish. To make this concrete, without any other considered criteria, the matrices (L, A^c, B^c) leading to the smallest $\mathcal{E}^x_{\Gamma,k}$ are trivially given by $(L, A^c, B^c) = \mathbf{0}$. While this is effective at eliminating the impact of the attacker, it implies that we discard the observer and the controller altogether and, therefore, forfeit any ability to control the system and build a reliable estimate of the state. If there are performance specifications that the observer, monitor, and controller must satisfy in the attack-free case (e.g., convergence speed, perturbation-output gain, and closed-loop dynamics spectrum), they have to be added as extra constraints into the minimization problem posted to minimize the volume of $\mathcal{E}^x_{\Gamma,\infty}$.

Several time and frequency domain performance specifications for LTI systems have been expressed as LMI constraints on the closed-loop state-space matrices and quadratic Lyapunov functions [31]. Here, our goal is to compute a single observer (12), monitor (16), and controller (17) that: 1) meets the required attack-free performance specifications, and 2) decreases the set of states reachable by stealthy attackers. For LTI systems and some of the most frequently used performance specifications (e.g., general quadratic performance [31]), there are analysis and synthesis results of the form: System Σ satisfies the performance specification γ_j if there exists a Lyapunov matrix $\mathcal{P}_j$ that satisfies some LMIs in $\mathcal{P}_j$. If our synthesis problem involves N specifications, $\gamma_1, \dots, \gamma_N$, by collecting the LMIs of each specification, we end up having a set of matrix inequalities whose variables are the observer, monitor, and controller matrices, and the Lyapunov matrices, $\mathcal{P}_1, \dots, \mathcal{P}_N$, of the specifications (plus auxiliary variables depending on the performance criteria). To pose a tractable co-design considering the volume of $\mathcal{E}^x_{\Gamma,k}$ and the specification γ_j, we must rewrite the specification Lyapunov matrix $\mathcal{P}_j$ and its corresponding LMIs in terms of the synthesis variables ν. This can be achieved by imposing $\mathcal{P}_j = T_j^T \mathcal{P} T_j$, where $\mathcal{P}$ is the Lyapunov-like matrix associated with $\mathcal{E}^x_{\Gamma,k}$ in (33) and T_j denotes some linear transformation. By doing so, we can write the specification LMIs in terms of $\mathcal{P}$ and use the change of variables in (40) and the transformations $\mathcal{T}_1$ and $\mathcal{T}_2$ in (35)-(36) to write these LMIs in terms of ν.

Remark 7 *In this manuscript, as attack-free performance specifications, we consider the spectrum of the estimation error dynamics for the observer and, for the controller, the $\mathcal{L}_2$ gain from the vector of perturbations to some performance output. We remark that any other specification γ_j can be considered in our framework as long as the corresponding Lyapunov matrix $\mathcal{P}_j$ and the LMIs can be written in terms of the synthesis variables ν. In Ref. [31], the authors provide a synthesis framework for general quadratic performance – which covers $\mathcal{H}_2/\mathcal{H}_\infty$ performance, passivity, asymptotic disturbance rejection, peak impulse response, peak-to-peak gain, nominal/robust regulation, and closed-loop pole location. The framework here and the one in [31] are compatible in the sense that any performance specification considered in [31] can be written as LMIs in terms of our syntheses variables ν.*

Attack-Free Monitor Feasibility. Note that the observer gain L and the monitor matrix Π must be chosen such that Assumption 1 is satisfied. That is, the pair (L, Π) must be selected such that, in the attack-free case ($\delta_k = \mathbf{0}$), there exists some $k^* \in \mathbb{N}$ satisfying $r_k^T \Pi r_k \leq 1$ for all $k \geq k^*$ and r_k solution of (15). Next, we provide constraints in the syntheses variables ν that have to be fulfilled to satisfy Assumption 1.

Lemma 5 *Consider the system matrices (A^p, C^p, E, F) and the perturbation bounds $\bar{v}, \bar{\eta} \in \mathbb{R}_{>0}$. Assume no attacks to the system, i.e., $\delta_k = \mathbf{0}$. For a given $a \in (0,1)$, constant $\alpha^e_\infty := (2-a)/(1-a)$, and $\epsilon \in \mathbb{R}_{>0}$, if there exist constants $a_1, a_2 \in \mathbb{R}$ and matrices $S \in \mathbb{R}^{n\times n}$, $G \in \mathbb{R}^{m\times m}$, and $R \in \mathbb{R}^{n\times m}$ satisfying:*

$$\begin{cases} a_1, a_2 \in (0,1),\ a_1 + a_2 \geq a,\ S > \mathbf{0},\ G > \mathbf{0}, \\ \begin{bmatrix} aS & (SA^p - RC^p)^T & \mathbf{0} & \mathbf{0} \\ SA^p - RC^p & S & -RF & SE \\ \mathbf{0} & -(RF)^T & \frac{1-a_1}{\bar{\eta}} I_m & \mathbf{0} \\ \mathbf{0} & E^T S & \mathbf{0} & \frac{1-a_2}{\bar{v}} I_n \end{bmatrix} \geq \mathbf{0}, \\ \begin{bmatrix} \frac{1}{\alpha^e_\infty + \epsilon + \bar{\eta}} S - (C^p)^T G C^p & -(C^p)^T G \\ -GC^p & \frac{1}{\alpha^e_\infty + \epsilon + \bar{\eta}} I_m - G \end{bmatrix} \geq \mathbf{0}; \end{cases} \tag{48}$$

then, for $L = S^{-1}R$ and $\Pi = G$, the residual dynamics (15) satisfies $r_k^T \Pi r_k \leq 1$ for all $k \geq k^*(a, \epsilon, e_1, S)$, where $k^*(a, \epsilon, e_1, S) := \min\{k \in \mathbb{N} | a^{k-1}\left(e_1^T S e_1 - \alpha^e_\infty\right) \leq \epsilon\}$ and e_1 denotes the initial estimation error in (15).

The proof of Lemma 5 is given in the appendix. The constant ϵ determines the tightness of the monitor, i.e., the smaller the ϵ the tighter the bound $r_k^T \Pi r_k \leq 1$ for $k \geq k^*$. Note, however, that depending on the initial condition e_1, too small ϵ might result in very large $k^* = \min\{k \in \mathbb{N} | a^{k-1}\left(e_1^T S e_1 - \alpha^e_\infty\right) \leq \epsilon\}$. See Remark 11 in the appendix for further details.

Attack-Free Observer Performance. For the observer, we simply consider the speed of convergence of the estimation error to steady state as a performance criteria. This is quantified by the eigenvalues of the matrix $(A^p - LC^p)$. We restrict the values that L might take by enforcing that the eigenvalues of $(A^p - LC^p)$ are contained in a disk, Disk$[\beta, \tau]$, centered at $\beta + 0i$ with radius τ. We give a necessary and sufficient condition in terms of the synthesis variables, R and S, to achieve this performance.

Lemma 6 [Observer Performance][10] *Consider the system matrices (A^p, C^p). If there exist $S \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{n\times m}$ satisfying:*

$$\begin{cases} S > 0, \\ \begin{bmatrix} S & \beta S - SA^p + RC^p \\ (\alpha S - SA^p + RC^p)^T & \tau^2 S \end{bmatrix} \geq \mathbf{0}; \end{cases} \tag{49}$$

then, the eigenvalues of $(A^p - LC^p)$ with $L = S^{-1}R$ are contained in the closed disk Disk$[\beta, \tau]$ centered at $\beta + 0i$ with radius τ.

Attack-Free Controller Performance. For the controller, we consider the $\mathcal{L}_2$ gain of the closed-loop system from the vector of perturbations, $d_k := (\eta_k^T, v_k^T)^T \in \mathbb{R}^{m+n}$, to some performance output, say $s_k \in \mathbb{R}^g$, in the attack-free case (i.e., $\delta_k = \mathbf{0}$). Define the matrices

$$\tilde{\mathcal{A}} := \begin{bmatrix} A^p + B^p D^c C^p & B^p C^c \\ B^c C^p & A^c \end{bmatrix}, \ \tilde{\mathcal{B}} := \begin{bmatrix} B^p D^c F & E \\ B^c F & \mathbf{0} \end{bmatrix}, \tag{50}$$

and the performance output $s_k := C^s x_k^p + D^s u_k + D_1 \eta_k + D_2 v_k$, for some matrices $C^s \in \mathbb{R}^{g\times n}$, $D^s \in \mathbb{R}^{g\times l}$, $D_1 \in \mathbb{R}^{g\times m}$, and $D_2 \in \mathbb{R}^{g\times n}$. Then, the closed-loop dynamics (11),(17) can be written in terms of the extended state $\tilde{\zeta}_k := ((x_k^p)^T, (x_k^c)^T)^T \in \mathbb{R}^{2n}$, the vector of perturbations d_k, and the performance output s_k:

$$\begin{cases} \tilde{\zeta}_{k+1} = \tilde{\mathcal{A}}\tilde{\zeta}_k + \tilde{\mathcal{B}} d_k, \\ s_k = \tilde{\mathcal{C}}\tilde{\zeta}_k + \tilde{\mathcal{D}} d_k, \end{cases} \tag{51}$$

with $\tilde{\mathcal{C}} := (C^s + D^s D^c C^p, D^s C^c)$ and $\tilde{\mathcal{D}} := (D_1 + D^s D^c F, D_2)$. The $\mathcal{L}_2$ gain from d_k to s_k of system (51) is given by $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} (\|s_k\|_2 / \|d_k\|_2)$ for $\tilde{\zeta}_1 = \mathbf{0}$, where, for any sequence $\rho_k \in \mathbb{R}^{n_\rho}$, $\|\rho_k\|_2 := \sum_{k=1}^{\infty} (\rho_k^T \rho_k)^{\frac{1}{2}}$. The $\mathcal{L}_2$ gain of system (51) equals the $\mathcal{H}_\infty$ norm of the transfer matrix $H(s) := \tilde{\mathcal{D}} + \tilde{\mathcal{C}}(sI - \tilde{\mathcal{A}})^{-1}\tilde{\mathcal{B}}$, see [32].

Lemma 7 [Bounded-Real Lemma] *Consider the closed-loop system (51) with input d_k and output s_k. If there exist $\mathcal{X} \in \mathbb{R}^{2n\times 2n}$ and $\gamma \in \mathbb{R}_{>0}$ satisfying:*

$$\mathcal{X} > \mathbf{0}, \ \ \mathcal{S} := \begin{bmatrix} \mathcal{X} & \tilde{\mathcal{A}}^T\mathcal{X} & \mathbf{0} & \tilde{\mathcal{C}}^T \\ \mathcal{X}\tilde{\mathcal{A}} & \mathcal{X} & \mathcal{X}\tilde{\mathcal{B}} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathcal{B}}^T\mathcal{X} & \gamma^2 I & \tilde{\mathcal{D}}^T \\ \tilde{\mathcal{C}} & \mathbf{0} & \tilde{\mathcal{D}} & I \end{bmatrix} \geq \mathbf{0}; \tag{52}$$

then, the $\mathcal{L}_2$ gain of system (51) is less than or equal to γ, i.e., $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} (\|s_k\|_2 / \|d_k\|_2) \leq \gamma$ for $\tilde{\zeta}_1 = \mathbf{0}$.

The proof of Lemma 7 is omitted here. It is a standard result and details about the proof can be found in, for instance, [4], [32], and references therein.

Using the analysis inequalities in (52), we derive the corresponding synthesis constraints in terms of the synthesis variables ν. Consider the matrices $\mathcal{X}$ and $\mathcal{Y}$ introduced in (34), the change of variables in (40a), and the attack-free closed-loop system matrices $(\tilde{\mathcal{A}}, \tilde{\mathcal{B}}, \tilde{\mathcal{C}}, \tilde{\mathcal{D}})$ above defined. Define the matrices:

$$\begin{cases} \tilde{\mathbf{X}}(\nu) := \mathcal{Y}^T\mathcal{X}\mathcal{Y} = \begin{bmatrix} Y & I \\ I & X \end{bmatrix}, \\ \tilde{\mathbf{A}}(\nu) := \mathcal{Y}^T\mathcal{X}\tilde{\mathcal{A}}\mathcal{Y} = \begin{bmatrix} A^pY + B^pM & A^p + B^pNC^p \\ K & XA^p + OC^p \end{bmatrix}, \\ \tilde{\mathbf{B}}(\nu) := \mathcal{Y}^T\mathcal{X}\tilde{\mathcal{B}} = \begin{bmatrix} B^pNF & E \\ OF & XE \end{bmatrix}, \\ \tilde{\mathbf{C}}(\nu) := \tilde{\mathcal{C}}\mathcal{Y} = [C^sY + D^sM \quad C^s + D^sNC^p], \\ \tilde{\mathbf{D}}(\nu) := \tilde{\mathcal{D}} = [D_1 + D^sNF \quad D_2]. \end{cases} \tag{53}$$

Lemma 8 [$\mathcal{H}_\infty$-Performance] *Consider the system matrices (A^p, B^p, C^p, E, F). If there exist $O \in \mathbb{R}^{n\times l}$, $X, Y, K \in \mathbb{R}^{n\times n}$, $M \in \mathbb{R}^{m\times n}$, and $N \in \mathbb{R}^{l\times m}$, and constant $\gamma \in \mathbb{R}_{>0}$ satisfying:*

$$\tilde{\mathbf{X}}(\nu) > \mathbf{0}, \ \ \mathbf{S}(\nu) := \begin{bmatrix} \tilde{\mathbf{X}}(\nu) & \tilde{\mathbf{A}}(\nu)^T & \mathbf{0} & \tilde{\mathbf{C}}(\nu)^T \\ \tilde{\mathbf{A}}(\nu) & \tilde{\mathbf{X}}(\nu) & \tilde{\mathbf{B}}(\nu) & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{B}}(\nu)^T & \gamma^2 I & \tilde{\mathbf{D}}(\nu)^T \\ \tilde{\mathbf{C}}(\nu) & \mathbf{0} & \tilde{\mathbf{D}}(\nu) & I \end{bmatrix} \geq \mathbf{0}; \tag{54}$$

then, the change of variables in (40a) and the matrix $\mathcal{Y}$ in (34) are invertible and the matrices $(\mathcal{X}, A^c, B^c, C^c, D^c)$ obtained by inverting (40a) and $\tilde{\mathbf{X}}(\nu) = \mathcal{Y}^T\mathcal{X}\mathcal{Y}$ in (53) satisfy (52) and lead to $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} \frac{\|s_k\|_2}{\|d_k\|_2} \leq \gamma$ for $\tilde{\zeta}_1 = \mathbf{0}$.

The proof of Lemma 8 is given in the appendix.

5.3 Synthesis of Secure Control Systems

Finally, combining the results above presented, we cast the complete optimization problem to minimize the volume of the asymptotic approximation $\mathcal{E}^x_{\Gamma,\infty}$ of $\mathcal{R}^x_{\Gamma,k}$ as a function of the set of sensor being attacked (the sensor selection matrix Γ) while guaranteing certain attack-free system performance.

Theorem 2 *Consider (A^p, B^p, C^p, E, F) (the system matrices), the perturbations bounds $\bar{v}, \bar{\eta} \in \mathbb{R}_{>0}$, and the attack sensor selection matrix Γ. For given $a, b \in (0,1)$, $\alpha^e_\infty = (2-a)/(1-a)$, $\epsilon \in \mathbb{R}_{>0}$, $\tau, \beta \in (0,1)$, $\gamma \in \mathbb{R}_{>0}$, if there exist $a_1, a_2 \in \mathbb{R}$ and matrices $\nu = (X, Y, S, R, G, K, O, M, N)$, $X, Y, S, K \in \mathbb{R}^{n\times n}$, $R \in \mathbb{R}^{n\times m}$, $G \in \mathbb{R}^{m\times m}$, $O \in \mathbb{R}^{n\times l}$, $M \in \mathbb{R}^{m\times n}$, $N \in \mathbb{R}^{l\times m}$, solution of the convex optimization:*

$$\min_{\nu, a_1, a_2} \ \text{trace}[Y], \tag{55a}$$

$$\begin{cases} \text{s.t. } a_1, a_2 \in (0,1), \ a_1 + a_2 + b \geq a, \\ \mathbf{P}(\nu) > \mathbf{0}, \ \mathbf{L}(\nu) \geq \mathbf{0}, \ \text{(attacker's reachable set)}, \\ (48), \ \text{(monitor feasibility)}, \\ (49), \ \text{(observer performance)}, \\ \tilde{\mathbf{X}}(\nu) > \mathbf{0}, \ \mathbf{S}(\nu) \geq \mathbf{0}, \ \text{(controller performance)}, \end{cases} \tag{55b}$$

with $\mathbf{P}(\nu), \mathbf{L}(\nu), \tilde{\mathbf{X}}(\nu)$, and $\mathbf{S}(\nu)$ as defined in (37), (46), (53), and (54), respectively; then, the transformation $\mathcal{T}_1$ in (35) and the change of variables in (40) are invert-

ible and the matrices $(\mathcal{P}, L, \Pi, A^c, B^c, C^c, D^c)$ *obtained by inverting* (40) *and* $\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 = \mathbf{P}(\nu)$ *in* (37) *lead to: 1) a feasible monitor in the sense of Lemma 5; 2) for* $k \geq k^*(a, \epsilon, e_1, S) = \min\{k \in \mathbb{N} | a^{k-1}\big(e_1^T S e_1 - \alpha_\infty^e\big) \leq \epsilon\}$ *and initial estimation error* e_1 *in* (15), $\mathcal{R}_{\Gamma,k}^x \subseteq \mathcal{E}_{\Gamma,k}^x$ *with* $\mathcal{E}_{\Gamma,k}^x = \{x^p \in \mathbb{R}^{3n} | (x^p)^T \mathcal{P}_\Gamma^x x^p \leq \alpha_k^\zeta\}$, $\mathcal{P}_\Gamma^x := X - U\tilde{X}^{-1}U^T$, *and* $\alpha_k^\zeta := a^{k-1}\zeta_{k^*}^T \mathcal{P} \zeta_{k^*} + \frac{3-a}{1-a}(1-a^{k-1})$; *3) the eigenvalues of* $(A^p - LC^p)$ *being contained in* $\text{Disk}[\beta, \tau]$; *and 4)* $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} (\|s_k\|_2 / \|d_k\|_2) \leq \gamma$ *for* $\tilde{\zeta}_1 = \mathbf{0}$. *Moreover, by minimizing* trace$[Y]$, *we are minimizing an upper bound on the volume of* $\mathcal{E}_{\Gamma,\infty}^x$.

Proof: Assume that (ν, a_1, a_2) satisfy the constraints in (55). By Lemma 2, because $\mathbf{P}(\nu) > \mathbf{0}$ and $\mathbf{L}(\nu) \geq \mathbf{0}$, the transformation $\mathcal{T}_1$ in (35) and the change of variables in (40) are invertible, and the $(\mathcal{P}, L, \Pi, A^c, B^c, C^c, D^c)$ obtained by inverting (40) and $\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 = \mathbf{P}(\nu)$ in (37) satisfy the analysis inequalities $\mathcal{P} > \mathbf{0}$ and $\mathcal{L} \geq \mathbf{0}$ defined in (31) and (33), respectively, and are unique. Moreover, by assumption, (48) is fulfilled. Then, by Lemma 5, the residual dynamics (15) satisfies $r_k^T \Pi r_k \leq 1$ for all $k \geq k^*(a, \epsilon, e_1, S)$, and $\Pi = G$ and $L = S^{-1}R$. Therefore, by Lemma 2, Lemma 3, and Lemma 5, $\mathcal{R}_{\Gamma,k}^x \subseteq \mathcal{E}_{\Gamma,k}^x$ with $\mathcal{P}_\Gamma^x = X - U\tilde{X}^{-1}U^T$ and $\alpha_k^\zeta = a^{k-1}\zeta_{k^*}^T \mathcal{P} \zeta_{k^*} + \frac{3-a}{1-a}(1-a^{k-1})$. Because we are minimizing trace$[Y]$ and $Y = (X - U\tilde{X}^{-1}U^T)^{-1}$, by Lemma 3 and Lemma 4, we are minimizing an upper bound on the volume of $\mathcal{E}_{\Gamma,\infty}^x$. Next, because (49) is fulfilled by assumption, by Lemma 6, the eigenvalues of $(A^p - LC^p)$ with $L = S^{-1}R$ are contained in $\text{Disk}[\beta, \tau]$. Finally, because ν satisfy $\tilde{\mathbf{X}}(\nu) > \mathbf{0}$ and $\mathbf{S}(\nu) \geq \mathbf{0}$ by assumption, by Lemma 8, the controller obtained by inverting (40a) leads to $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} (\|s_k\|_2 / \|d_k\|_2) \leq \gamma$. ■

Observer, Monitor, Controller, and Ellipsoidal-Approximation Reconstruction. Given a solution (ν, a_1, a_2) of the optimization problem in (55):

(1) For given X and Y, compute via singular value decomposition a full rank factorization $VU^T = I - YX$ with square and nonsingular V and U.

(2) For given ν and invertible V and U, solve the system of equations $\mathcal{T}_1^T \mathcal{P} \mathcal{T}_1 = \mathbf{P}(\nu)$ and (40) to obtain the matrices $(\mathcal{P}, L, \Pi, A^c, B^c, C^c, D^c)$.

(3) For given S, Y, $\mathcal{P}$, e_1, ϵ, and a, obtain the monitor convergence time k^*, and $\mathcal{P}_\Gamma^x$ and α_k^ζ conforming the ellipsoidal approximation $\mathcal{E}_{\Gamma,k}^x$ of $\mathcal{R}_{\Gamma,k}^x$ as: $k^* = \min\{k \in \mathbb{N} | a^{k-1}\big(e_1^T S e_1 - \alpha_\infty^e\big)$, $\mathcal{P}_\Gamma^x = Y^{-1}$, and $\alpha_k^\zeta = a^{k-1}\zeta_{k^*}^T \mathcal{P} \zeta_{k^*} + \frac{3-a}{1-a}(1-a^{k-1})$.

By Theorem 2, the reconstructed matrices satisfy the attack-free system performance, and minimize an upper bound on the volume of $\mathcal{E}_{\Gamma,\infty}^x$.

Remark 8 *To obtain tighter approximations* $\mathcal{E}_{\Gamma,k}^x$ *of* $\mathcal{R}_{\Gamma,k}^x$, *once the matrices* $(L, \Pi, A^c, B^c, C^c, D^c)$ *are computed using Theorem 2 and the above reconstruction procedure, we can close the loop using these matrices and use the analysis result in Theorem 1 to obtain tighter approximations. That is, Theorem 2 could be used for synthesis only, and then, once* $(L, \Pi, A^c, B^c, C^c, D^c)$ *are computed, we could use the analysis result in Theorem 1 to obtain less conservative approximations of* $\mathcal{R}_{\Gamma,k}^x$.

Remark 9 *Note that the constants* $a, b, \epsilon, \tau, \beta$, *and* γ *in Theorem 2 must be fixed before solving the synthesis optimization problem in* (55). *The constants* (τ, β, γ) *determine the attack-free observer and controller performance. The constant* ϵ *determines the tightness of the monitor in the attack-free case. The smaller the* ϵ *the tighter the monitor (see Remark 11 in the Appendix for details). Finally,* $a, b \in (0, 1)$ *are, in fact, variables of the optimization problem. However, to linearize some of the constraints, we fix their value before solving* (55) *and search over* $a, b \in (0, 1)$ *to find the optimal* ν. *The latter increases the computations needed to find the optimal* ν; *however, because* $a, b \in (0, 1)$ *(a bounded set), the required grid in* (a, b) *is of reasonable size.*

5.4 Distance to Critical States: Synthesis

As a second cost function for synthesis, we consider the distance between $\mathcal{R}_{\Gamma,k}^x$ and a possible set of critical states $\mathcal{C}^x$. Because $\mathcal{R}_{\Gamma,k}^x$ is not known exactly, we consider the distance from the approximation $\mathcal{E}_{\Gamma,k}^x$ to $\mathcal{C}^x$ and use this distance as cost function. We capture the set of critical states through the union of half-spaces defined by their boundary hyperplanes as introduced in (28). In the analysis case, we compute the *minimum distance*, $d_{\Gamma,k}^x$, between $\mathcal{E}_{\Gamma,k}^x$ and $\mathcal{C}^x$ and use this distance to approximate the proposed security metric (the distance between $\mathcal{R}_{\Gamma,k}^x$ and $\mathcal{C}^x$). For synthesis, however, the distance $d_{\Gamma,k}^x$ is highly nonlinear and not convex/concave in the syntheses variables ν. Instead, we consider the minimum distance between each hyperplane conforming $\mathcal{C}^x$ and the asymptotic ellipsoidal approximation, $\mathcal{E}_{\Gamma,\infty}^x = \lim_{k\to\infty} \mathcal{E}_{\Gamma,k}^x$, and use the weighted sum of these distances as the cost function to be maximized.

Proposition 2 *Consider the ellipsoidal approximation* $\mathcal{E}_{\Gamma,k}^x$ *as introduced in Lemma 3 with matrix* Y *and function* α_k^ζ, *and the set of critical states:*

$$\mathcal{C}^x = \left\{ x^p \in \mathbb{R}^n \;\middle|\; \bigcup_{i=1}^N c_i^T x^p \geq b_i \right\},$$

where each pair (c_i, b_i), $c_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \ldots, N$ *quantifies a hyperplane that defines a single half-space. The minimum distance* $d_{\Gamma,k}^{x,i}$ *between* $\mathcal{E}_{\Gamma,k}^x$ *and the hyperplane* $c_i^T x^p = b_i$ *is given by* $d_{\Gamma,k}^{x,i} := \frac{|b_i| - \sqrt{c_i^T Y c_i / \alpha_k^\zeta}}{c_i^T c_i}$.

Proof: The assertion follows by the same arguments as in the proof of Corollary 3. ■

For synthesis, we aim at maximizing $\sum_{i=1}^N \rho_i d_{\Gamma,k}^{x,i}$, for some $\rho_i \in \mathbb{R}_{\geq 0}$ satisfying $\sum_{i=1}^N \rho_i = 1$, by selecting

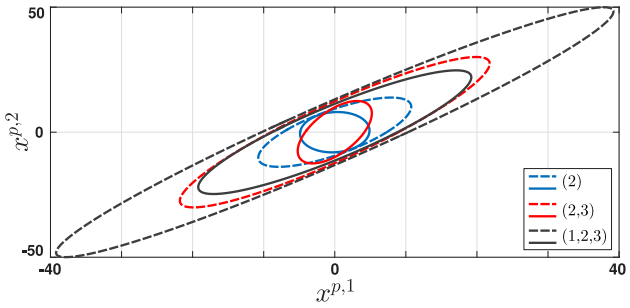

Fig. 7. Projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,1}, x^{p,2})$-hyperplane for different sets of sensor being attacked and distance to critical states. Continuous-lines correspond to the original κ in (30) and dashed-lines to the optimal κ obtained using Theorem 2.

(ν, a_1, a_2) subject to (55b). The constant ρ_i assigns a priority weight to the distance $d^{x,i}_{\Gamma,k}$. Note, however, that because $\alpha^\zeta_k = a^{k-1}\zeta^T_{k^*}\mathcal{P}\zeta_{k^*} + \frac{3-a}{1-a}(1-a^{k-1})$ and

$$\mathcal{P}^{-1} = \text{diag}\left[\begin{pmatrix} Y & V \\ V^T & \tilde{Y} \end{pmatrix}, S^{-1}\right],$$

the term $c_i^T Y c_i/\alpha^\zeta_k$ is nonlinear and not convex/concave in the matrix Y. However, because $a \in (0,1)$, we can maximize the weighted sum of the asymptotic minimum distances between $\mathcal{E}^x_{\Gamma,k}$ and $c_i^T x^p = b_i$, $i = 1,\dots,N$, i.e., $\tilde{d}_\Gamma := \lim_{k\to\infty} \sum_{i=1}^N \rho_i d^{x,i}_{\Gamma,k} = \sum_{i=1}^N \rho_i\big(|b_i| - (\frac{1-a}{3-a}c_i^T Y c_i)^{-1/2}\big)/c_i^T c_i$. Because $(1-a)/(3-a)$ is strictly positive and Y is positive definite, maximizing $\tilde{d}_\Gamma$ is equivalent to minimizing the linear function: $\sum_{i=1}^N \rho_i(c_i^T Y c_i)$. Next, as a corollary of Theorem 2, we pose the optimization problem required to maximize $\tilde{d}_\Gamma$ while guaranteeing the required attack-free performance.

Corollary 4 *Consider the setting stated in Theorem 2, the set of critical states $\mathcal{C}^x$ defined in (28), and $\tilde{d}_\Gamma$ above defined for some $\rho_i \in \mathbb{R}_{\geq 0}$, $\sum_{i=1}^N \rho_i = 1$, $i = 1,\dots,N$. If there exists (ν, a_1, a_2) solution of the optimization:*

$$\begin{cases} \min_{\nu,a_1,a_2} \ \sum_{i=1}^N \rho_i c_i^T Y c_i, \\ \text{s.t. (55b)}, \end{cases} \tag{56}$$

then, the matrices $(\mathcal{P}, L, \Pi, A^c, B^c, C^c, D^c)$ obtained by inverting (40) and $\mathcal{T}_1^T\mathcal{P}\mathcal{T}_1 = \mathbf{P}(\nu)$ in (37), maximize $\tilde{d}_\Gamma$ and satisfy the desired attack-free system performance in the sense of Theorem 2.

Proof: Let the constraints in (55b) be satisfied. Then, by the same arguments as stated in the proof of Theorem 2, the matrices $(\mathcal{P}, L, \Pi, A^c, B^c, C^c, D^c)$ obtained by inverting (40) and $\mathcal{T}_1^T\mathcal{P}\mathcal{T}_1 = \mathbf{P}(\nu)$ in (37) lead to a closed-loop dynamics that satisfies the attack-free performance considered in Theorem 2. Also, by the arguments above presented, minimizing $\sum_{i=1}^N \rho_i c_i^T Y c_i$ and maximizing $\tilde{d}_\Gamma$ are equivalent objectives. ■

5.5 *Controller/Monitor Selection for Unknown* Γ

The synthesis results presented above are derived for *given* sensor attack selection matrix Γ, see Remark 6. However, we do not have access to Γ in practice, i.e., the set of sensors being attacked is usually unknown to the system designer. Next, we provide general guidelines for using the results given above to synthesize controllers/monitors when Γ is unknown. We propose two sets of techniques: *sensor protection placement methods* [9,15]; and *game-theoretic techniques* [3].

Sensor Protection Placement. This technique was originally introduced for power system [9,15]. The problem is the following: assuming that the system designer has *limited security resources* to completely encrypt and secure a subset of sensors (i.e., attacks to those sensors are impossible), how to select which sensors to secure in order to minimize the effect of stealthy attacks on the system performance. In exactly the same sense, we have shown in the analysis example in Section 4.3 that our analysis tools can be used to allocate security equipment to sensors when limited resources are available so that the size of the stealthy reachable set is minimized. Now, in the syntheses setting, we address a slightly different problem: for given m sensors and a limited number of sensors that can be secured $\tilde{m} \in \{1,\dots,m\}$, which sensors should be selected such that the optimal controller/monitor corresponding to attacks to all the remaining $m-\tilde{m}$ sensors leads to the smallest stealthy reachable set (or the largest distance to critical states) among all subsets of $m-\tilde{m}$ sensors. For instance, assume that we have three sensors, $m = 3$, and $\tilde{m} = 1$ of them can be secured. Then, among all subsets of sensors $J \subseteq \{1,2,3\}$ with cardinality $\text{card}[J] = m - \tilde{m} = 2$ (i.e., $J \in \{\{1,2\},\{1,3\},\{2,3\}\}$), select the controller/monitor $\kappa_J \in \{\kappa_{\{1,2\}}, \kappa_{\{1,3\}}, \kappa_{\{2,3\}}\}$ that leads to the smallest asymptotic ellipsoid $\mathcal{E}^x_J := \mathcal{E}^x_{\Gamma,\infty}|_{\Gamma=\Gamma_J,\kappa=\kappa_J}$, where κ_J denotes the optimal $\kappa = (L, \Pi, A^c, B^c, C^c, D^c)$ corresponding to the solution of (55) for $\Gamma = \Gamma_J$, and $\Gamma_J \in \{\Gamma_{\{1,2\}}, \Gamma_{\{1,3\}}, \Gamma_{\{2,3\}}\}$ is the attack selection matrix corresponding to attacks on sensors J. That is, we compute optimal controllers/monitors and corresponding asymptotic ellipsoids $(\kappa_J, \mathcal{E}^x_J)$ for all $J \in \{\{1,2\},\{1,3\},\{2,3\}\}$, and select the controller κ_J that leads to the smallest $\mathcal{E}^x_J$. In the following algorithm, we summarize the ideas introduced above.

Algorithm 1. Controller/Monitor Selection:

1) Consider the m available sensors, the number of sensors that can be secured $\tilde{m} \in \{1,\dots,m\}$, and all subsets of sensors $J \subseteq \{1,\dots,m\}$ with cardinality $\text{card}[J] = m - \tilde{m}$.

2) Let Γ_J denote the sensor attack selection matrix corresponding to attacks on sensors J. For $\Gamma = \Gamma_J$ and all $J \subseteq \{1,\dots,m\}$ with $\text{card}[J] = m-\tilde{m}$, compute the optimal controller/monitor $\kappa_J := \kappa = (L, \Pi, A^c, B^c, C^c, D^c)$ corresponding to the solution of (55) in Theorem 2.

3) Let $\mathcal{E}^x_J = \mathcal{E}^x_{\Gamma,\infty}|_{\Gamma=\Gamma_J,\kappa=\kappa_J}$, i.e., $\mathcal{E}^x_J$ denotes the asymptotic ellipsoidal approximation of the stealthy reachable set, $\mathcal{R}^x_{\Gamma,k}$, for $\Gamma = \Gamma_J$ and $\kappa = \kappa_J$; and select the controller/monitor as follows:

$$\kappa^*_{\tilde{m}} = \arg\min_{\kappa_J} \text{Vol}[\mathcal{E}^x_J], \tag{57}$$

where $\text{Vol}[\mathcal{E}^x_J]$ denotes the volume of $\mathcal{E}^x_J$.

Note that the selected controller/monitor $\kappa^*_{\tilde{m}}$ in (57) is parametrized by $\tilde{m}$, the number of sensors that can be secured; and that in the case $\tilde{m} = 0$ (no sensors can be secured), $\Gamma_J = \Gamma_{\{1,\dots,m\}} = I_m$, i.e., the selected controller/monitor κ^*_0 is a worst-case controller that assumes all sensors are attacked. We remark that Algorithm 1 could be used using the largest distance to critical states $\tilde{d}_\Gamma$ as cost to be *maximized* instead of minimizing $\text{Vol}[\mathcal{E}^x_J]$.

Game-Theoretic Strategies. We only briefly introduce a game-theoretic formulation and some techniques that could be used to select suitable controllers/monitors for unknown Γ. A rigourous game-theoretic formulation is beyond the scope of this paper and is left as future work. Note that we can compute optimal controllers/monitors for all possible combinations of Γ. If there are m sensors, there are $\bar{m} := \sum_{s=1}^{m} \binom{m}{s}$ possible matrices Γ. We index and order all these matrices in the $\bar{m}$-tuple $(\Gamma_{\{1\}}, \Gamma_{\{2\}}, \dots, \Gamma_{\{1,2\}}, \Gamma_{\{1,3\}}, \dots, \Gamma_{\{1,\dots,m\}}) =: \bar{\Gamma}$, and the corresponding optimal controllers/monitors κ in $(\kappa_{\{1\}}, \kappa_{\{2\}}, \dots, \kappa_{\{1,2\}}, \kappa_{\{1,3\}}, \dots, \kappa_{\{1,\dots,m\}}) =: \bar{\kappa}$ with $\text{card}[\bar{\Gamma}] = \text{card}[\bar{\kappa}] = \bar{m}$, where, as introduced above, for instance, $\kappa_{\{1,3\}}$ is the controller/monitor κ corresponding to the solution of (55) in Theorem 2 for $\Gamma = \Gamma_{\{1,3\}}$, and $\Gamma_{\{1,3\}}$ is the sensor selection matrix Γ corresponding to attacks on sensors $\{1,3\}$. Associated with every pair $(\kappa_I, \Gamma_J) \in \bar{\kappa} \times \bar{\Gamma}$, we introduce the corresponding cost $h_{I,J} := \text{Vol}[\mathcal{E}^x_{\Gamma,\infty}|_{\Gamma=\Gamma_J,\kappa=\kappa_I}]$, where $\text{Vol}[\mathcal{E}^x_{\Gamma,\infty}|_{\Gamma=\Gamma_J,\kappa=\kappa_I}]$ denotes the volume of $\mathcal{E}^x_{\Gamma,\infty}$ for $\Gamma = \Gamma_J$ and $\kappa = \kappa_I$. That is, $\mathcal{E}^x_{\Gamma,\infty}|_{\Gamma=\Gamma_J,\kappa=\kappa_I}$ is the asymptotic ellipsoidal approximation of the stealthy reachable set $\mathcal{R}^x_{\Gamma,k}$ for $\Gamma = \Gamma_J$ and $\kappa = \kappa_I$, and κ_I is the controller/monitor κ corresponding to the solution of (55) in Theorem 2 for $\Gamma = \Gamma_I$. Note that, by construction, κ_I minimizes the cost $h_{I,I}$ but is not optimal for $h_{I,J}$, $I \neq J$. Next, using the notation introduced above, we cast the controller/monitor selection as a *two players noncooperative zero-sum matrix game* [3], where player one (the defender) has *strategy set* $\bar{\kappa}$, player two (the attacker) has *strategy set* $\bar{\Gamma}$, and the cost matrix of the game is $H := \{h_{I,J}\} \in \mathbb{R}^{\bar{m}\times\bar{m}}$. Define the tuple $K := (\{1\}, \{2\}, \dots, \{1,2\}, \{1,3\}, \dots, \{1,\dots,m\})$ indexed as $K_1 = \{1\}$, $K_2 = \{2\}$, $K_{\bar{m}} = \{1,\dots,m\}$, and so on. Then, elements of the game matrix $H(i,j)$, $i,j \in \{1,\dots,\bar{m}\}$, correspond to h_{K_i,K_j}, i.e., there is a one-to-one correspondence between $H(i,j)$ and h_{K_i,K_j}. Hereafter, we only use entries $H(i,j)$, $i,j \in \{1,\dots,\bar{m}\}$, of the matrix game without making reference to the corresponding sets (K_i, K_j); indeed, the strategy of the defender associated with $H(i,j)$ is κ_{K_i}, and the one of the attacker is Γ_{K_j}. If the defender chooses strategy i (the i-th row of H) and the attacker the strategy j (the j-th column of H), the outcome of the game is $H(i,j)$. Here, the defender seeks to minimize the outcome of the game, while the attacker aims at maximizing it, both by independent decisions. Note that this game is only played once, the controller is selected before the system starts operating and it is not changed during the operation. Then, a reasonable strategy for the defender is to secure his losses against any (rational or irrational) behavior of the attacker [3]. Under this strategy, the defender selects the strategy $i^* \in \{1,\dots,\bar{m}\}$, the i^*-row of H, whose largest entry is no bigger than the largest entry of any other row. Therefore, if the defender chooses the i^*-th row as his strategy, where i^* satisfies the inequalities:

$$\bar{f}(H) := \max_j H(i^*, j) \leq \max_j H(i,j); \tag{58}$$

then, his looses are no greater than $\bar{f}$, which is referred in the literature as *the ceiling* of the defender or the *security level* for the defender's losses [3]. The strategy "row i^*" (the controller/monitor $\kappa_{K_{i^*}}$) that yields this security level is called *the security strategy* of the defender. For every matrix game H, the security level of the defender's losses is unique, and there exists at least one security strategy [3]. Using the security strategy $\kappa_{K_{i^*}}$ (where i^* satisfies (58)) as the selected controller/monitor is the best rational strategy that can be taken under the assumptions of the game (i.e., noncooperative, played only once, and independent decisions). Note that the attacker has also a security strategy that secures his gains against any strategy of the defender. However, whether he plays that strategy (or not) would not chance the security strategy of the defender. Here, we use the security strategy $\kappa_{K_{i^*}}$ described above as the selected controller/monitor.

An alternative formulation is to assign probabilities to every strategy of the attacker and select the defender's strategy that minimizes the *expected value of the game*. Define the vector of probabilities $p := (p_1, \dots, p_{\bar{m}})^T$, $p_j \geq 0$, $\sum_{j=1}^{\bar{m}} p_j = 1$, $j \in \{1,\dots,\bar{m}\}$, where p_j denotes the probability that the attacker uses strategy j (the j-th column of H). These probabilities have to be assigned by the defender given the system configuration. For instance, if sensors are geographically distributed (e.g., in power/water networks), some of them could be completely inaccessible and some others might be easier to reach/hack. Another option is to assign higher probabilities to attacks on single sensors than on groups of them. Simply because it might be easier to hack one sensor than more than one. Thus, the system designer has to assign smaller/larger probabilities to every sensor of the system. Note that for a given defender's strategy i, the value of the game is $H(i,1)$ with probability p_1, $H(i,2)$ with probability p_2, $H(i,\bar{m})$ with probability $p_{\bar{m}}$, and so

on. Then, for this i-th row strategy, the expected value of the game is given by $H(i,*)p$, where $H(i,*) \in \mathbb{R}^{1\times \bar{m}}$ denotes the i-th row of the game matrix H. The defender selects the strategy $i^* \in \{1,\dots,\bar{m}\}$, the i^*-row of H, that minimizes the expected value of the game, i.e.,

$$i^* = \arg\min_i H(i^*,*)p. \tag{59}$$

We use the strategy "row i^*" (the controller/monitor $\kappa_{K_{i^*}}$) as the chosen controller/monitor. Note that this strategy might lead to a better outcome of the game (for the defender) with certain "optimal" probability. However, there is also a nonzero probability of doing worse than with the deterministic formulation presented above. This might be a risk worth taking to improve the security of the system. We remark that the matrix game H could be constructed using the largest distance to critical states $\tilde{d}_\Gamma$ instead of $\mathrm{Vol}[\mathcal{E}^x_{\Gamma,\infty}]$. In that case, the defender seeks to *maximize* the distance and the attacker aims at *minimizing* it.

5.6 Simulation Results

Consider the system matrices (A^p, B^p, C^p, E, F) in (30), and the perturbation bounds $\bar{\eta} = \sqrt{\pi}$ and $\bar{v} = 1$. Let $\epsilon = 0.1$ and $(\beta,\tau) = (0, 0.99)$, i.e., the monitor constant ϵ is fixed to 0.1 and the eigenvalues of the observer closed-loop matrix $(A^p - LC^p)$ are required to be contained in the disk centered at $0+0i$ of radius 0.80, Disk$[0, 0.80]$. Consider the performance output matrices $C_s = (0, 0, 0.25)$, $D^s = \mathbf{0}_{1\times 2}$, $D_1 = (0, 0, 1)$, and $D_2 = \mathbf{0}_{1\times 3}$, and the set of critical states $\mathcal{C}^x = \{x^p \in \mathbb{R}^3 | x^{p,1} \leq -15\}$. The controller must guarantee, in the attack-free case, that the $\mathcal{L}_2$-gain from the vector of perturbations $d_k = (\eta_k^T, v_k^T)^T$ to $s_k = C_s x_k^p + D_s u_k + D_1\eta_k + D_2 v_k = 0.25 x_k^{p,3} + \eta_k^3$ is less than or equal to $\gamma = 3.0$ (as the controller given in (30) for the analysis section). We use Theorem 2 and Corollary 4 to obtain optimal $\kappa = (L, \Pi, A^c, B^c, C^c, D^c)$ minimizing $\mathcal{E}^x_{\Gamma,\infty}$ and maximizing $\tilde{d}_\Gamma$, respectively, for all possible combinations of sensors being attacked (all the possible sensor attack selection matrices Γ). Once we have these κ, we use the analysis results in Theorem 1 and Corollary 2 to obtain tighter approximations $\mathcal{E}^x_{\Gamma,k}$ of $\mathcal{R}^x_{\Gamma,k}$; and use these $\mathcal{E}^x_{\Gamma,k}$ to obtain tighter $\tilde{d}_\Gamma$. As in the analysis case, we have k-dependent approximations $\mathcal{E}^x_{\Gamma,k}$; however, because $a < 1$, the function α^x_k conforming $\mathcal{E}^x_{\Gamma,k}$ converge exponentially to $(3-a)/(1-a)$. Hence, in a few time steps, $\mathcal{E}^x_{\Gamma,k} \approx \mathcal{E}^x_{\Gamma,\infty} = \{x \in \mathbb{R}^n | x^T \mathcal{P}^x_\Gamma x \leq (3-a)/(1-a)\}$, and thus, $\mathcal{E}^x_{\Gamma,k} \approx \mathcal{E}^x_{\Gamma,\infty}$. We present $\mathcal{E}^x_{\Gamma,\infty}$ instead of the time-dependent $\mathcal{E}^x_{\Gamma,k}$. In Table 2, we present the volume of the asymptotic approximation $\mathcal{E}^x_{\Gamma,\infty}$ and the distance $\tilde{d}_\Gamma$ between $\mathcal{E}^x_{\Gamma,\infty}$ and the critical states $\mathcal{C}^x$ for all possible combinations of sensors being attacked. We show results for the original κ in (30); and for the optimal κ obtained using Theorem 2 and Corollary 4. Note that the improvement is remarkable using the optimal κ. To illustrate this improvement, in Figure 7, we show the projection of $\mathcal{E}^x_{\Gamma,\infty}$ onto the $(x^{p,1}, x^{p,2})$-hyperplane for sensors {2},{2,3}, and {1,2,3} being attacked. We depict the projections for both the original κ in (30) and the optimal one (minimizing trace$[Y]$). For sensor {2}, we have a 67% improvement in volume and 142% in distance; for {2,3}, 88% and 247%; and for {1,2,3}, 67% and 92%, respectively. Once we have all the optimal controllers/monitors and the corresponding costs in Table 2, we can use Algorithm 1 in Section 5.5 (*the sensor protection placement method*) to select the best κ given a number of sensors that can be completely secured $\tilde{m}$. If $\tilde{m} = 1$; then, according to Algorithm 1, sensor two should be the one to be secured because $\kappa = \kappa_{1,3}$ (the optimal controller/monitor assuming sensors $\{1, 3\}$ are attacked) leads to the smallest volume (137.44), see Table 2. On the other hand, if distance to critical states is more important, the selected controller/monitor should be $\kappa_{2,3}$ (i.e., securing sensor one) because it leads to the largest distance (9.74). Following the same logic, if two sensors can be secured, $\tilde{m} = 2$, they should be sensors two and three, in terms of volume, and sensors one and two, in terms of distance, i.e., we should select controllers/monitors κ_1 (minimum volume) and κ_3 (maximum distance), respectively. Next, following the game-theoretic formulation in Section 5.5, using Theorem 2 for to all possible combinations of Γ, we compute all optimal controllers/monitors and the corresponding volumes of the ellipsoidal outer approximations. We use these volumes to construct the matrix game H (given in Table 3) as introduced in Section 5.5. Note that some entries of H are hyphens. This indicates that the optimization problem used to compute the ellipsoidal approximation was not feasible for that combination of controller/monitor and Γ. From this H, using 58, it is easy to verify that the *security level* for the defender's losses is 1538.31 which corresponds to controller/monitor $\kappa_{\{1,2,3\}}$ (the *the security strategy of the defender*), see Table 3. That is, by selecting $\kappa_{\{1,2,3\}}$, we ensure having a worst-case volume of 1538.31 regardless of what sensors the attacker compromises. Finally, we assign probabilities to the strategies of the attacker, in the sense introduced in Section 5.5, and look for the controller/monitor that minimizes the expected value of the game. Using (59), it is easy to verify that, for the vector of probabilities $p = (0.4, 0.09, 0.3, 0.1, 0.1, 0.01, 0)^T$, the strategy that minimizes the expected value of the game is $\kappa_{\{1\}}$, see Table 3. This controller/monitor leads to $H(1,*)p = 1135.43$, which is the smallest for all $H(i,*)$, $i \in \{1,\dots,7\}$.

6 Conclusion

We have provided mathematical tools – in terms of LMIs – for *quantifying* the potential impact of sensor stealthy attacks on the system dynamics. In particular, we have given a result for computing *ellipsoidal outer approximations* on the set of states that stealthy attacks can

Attacked Sensors	Original κ		**Cost:** min[trace[Y]] Optimal κ		**Cost:** min[$c^T Y c$] Optimal κ	
	Volume	**Distance**	**Volume**	**Distance**	**Volume**	**Distance**
{1}	150.72	8.07	116.94	9.12	150.16	9.27
{2}	453.51	4.20	145.31	10.15	151.80	10.98
{3}	219.43	8.60	130.62	10.77	194.50	11.92
{1,2}	952.95	-2.38	456.06	5.15	487.79	5.17
{1,3}	279.50	6.85	137.44	9.23	186.75	9.29
{2,3}	2063.46	-6.67	235.72	9.74	222.52	9.83
{1,2,3}	4300.32	-23.01	1394.31	-1.88	1371.94	-1.69

Table 2
Volume of the approximation $\mathcal{E}^x_{\Gamma,\infty}$ of $\mathcal{R}^x_{\Gamma,\infty}$ and distance $\tilde{d}_\Gamma$ to the critical states $\mathcal{C}^x$ for different sensors being attacked. We show results for the original κ in (30) and for the optimal κ obtained using Theorem 2 and Corollary 4.

κ/Γ	$\Gamma_{\{1\}}$	$\Gamma_{\{2\}}$	$\Gamma_{\{3\}}$	$\Gamma_{\{1,2\}}$	$\Gamma_{\{1,3\}}$	$\Gamma_{\{2.3\}}$	$\Gamma_{\{1,2,3\}}$
$\kappa_{\{1\}}$	116.94	3188.01	514.36	3104.51	514.73	29297.07	29991.81
$\kappa_{\{2\}}$	4277.61	145.31	2728.73	4233.15	38489.72	2489.05	37986.64
$\kappa_{\{3\}}$	302.15	8681.16	130.62	65681.23	300.58	8783.15	68909.86
$\kappa_{\{1,2\}}$	440.51	473.35	14207.72	456.06	15029.64	15746.28	17830.39
$\kappa_{\{1,3\}}$	134.27	86602.67	134.62	94982.02	137.44	73890.58	97253.67
$\kappa_{\{2,3\}}$	-	227.83	227.85	-	74255.39	235.72	-
$\kappa_{\{1,2,3\}}$	1184.02	1529.77	1435.6	1346.72	1320.99	1538.51	1394.31

Table 3
Noncooperative zero-sum matrix game between the attacker and the defender as introduced in Section 5.5.

induce in the system. We have proposed to use the *volume* of these approximations and the distance to possible dangerous states as *security metrics* for NCSs. Then, for given sensor attack selection matrix Γ, we have provide synthesis tools (in terms of semidefinite programs) to redesign controllers and monitors such that the impact of stealthy attacks is minimized and the required attack-free system performance is guaranteed. Based on these synthesis results, we have provided general guidelines for selecting optimal controllers/monitors when Γ is unknown. In particular, we have proposed two sets of techniques: *sensor protection placement methods*; and *game-theoretic techniques*. We have presented extensive computer simulations to illustrate the performance of our results.

Acknowledgements

This work was partially supported by the Australian Research Council (ARC) under the Discovery Project DP170104099.

References

[1] C. Z. Bai, F. Pasqualetti, and V. Gupta. Security in stochastic control systems: Fundamental limitations and performance bounds. In *American Control Conference (ACC), 2015*, pages 195–200, 2015.

[2] Cheng-Zong Bai and V. Gupta. On kalman filtering in the presence of a compromised sensor: Fundamental performance bounds. In *American Control Conference (ACC), 2014*, pages 3029–3034, 2014.

[3] T. Basar and G. Olsder. *Dynamic Noncooperative Game Theory, 2nd Edition*. Society for Industrial and Applied Mathematics, 1998.

[4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15 of *Studies in Applied Mathematics*. SIAM, Philadelphia, PA, 1994.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, New York, NY, USA, 2004.

[6] A. Cárdenas, S. Amin, Z. Lin, Y. Huang, C. Huang, and S. Sastry. Attacks against process control systems: risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, pages 355–366, 2011.

[7] Alvaro Cardenas, Saurabh Amin, Bruno Sinopoli, Annarita Giani, Adrian Perrig, and Shankar Sastry. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-physical Systems Security*, 2009.

[8] Jie Chen and Ron J. Patton. *Robust model-based fault diagnosis for dynamic systems*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.

[9] G. Dan and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *2010 First IEEE International Conference on Smart Grid Communications*, pages 214–219, 2010.

[10] G. Garcia and J. Bernussou. Pole assignment for uncertain systems in a specified disk by state feedback. *IEEE Transactions on Automatic Control*, 40:184–190, 1995.

[11] Ziyang Guo, D Shi, Karl Henrik Johansson, and Ling Shi. Optimal linear cyber-attack on remote state estimation. *IEEE Transactions on Control of Network Systems*, PP(99):1–10, 2016.

[12] F. Gustafsson. *Adaptive filtering and change detection*. John Wiley and Sons, LTD, West Sussex, Chichester, England, 2000.

[13] Roger A. Horn and Charles R. Johnson. *Matrix analysis*. Cambridge University Press, New York, NY, USA, 2nd edition, 2012.

[14] Sahand Hadizadeh Kafash, Jairo Giraldo, Carlos Murguia, Alvaro A. Cardenas, and Justin Ruths. Constraining attacker

capabilities through actuator saturation. In *proceedings of the American Control Conference (ACC), 2018*, 2018.

[15] T. T. Kim and H. V. Poor. Strategic protection against data injection attacks on power grids. *IEEE Transactions on Smart Grid*, 2:326–333, 2011.

[16] A. B Kurzhanskii and Istvan Valyi. *Ellipsoidal calculus for estimation and control*. Laxenburg, Austria : IIASA ; Boston : Birkhauser Boston, 1997.

[17] A. A. Kurzhanskiy and P. Varaiya. Ellipsoidal toolbox (et). In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 1498–1503, 2006.

[18] C. Kwon, W. Liu, and I. Hwang. Security analysis for cyber-physical systems against stealthy deception attacks. In *American Control Conference (ACC), 2013*, pages 3344–3349, 2013.

[19] Elias Kyriakides and Marios M. Polycarpou, editors. *Intelligent monitoring, control, and security of critical infrastructure systems*, volume 565 of *Studies in Computational Intelligence*. Springer, 2015.

[20] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas. Coding sensor outputs for injection attacks detection. In *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, pages 5776–5781, 2014.

[21] J. Milošević, H. Sandberg, and K. H. Johansson. Estimating the impact of cyber-attack strategies for stochastic control systems. In *arXiv:1811.05410*, 2018.

[22] M. I. Müller, J. Milošević, H. Sandberg, and C. R. Rojas. A risk-theoretical approach to $\mathcal{H}_2$-optimal control under covert attacks. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 4553–4558, 2018.

[23] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli. False data injection attacks against state estimation in wireless sensor networks. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 5967–5972, 2010.

[24] Y. Mo and B. Sinopoli. On the performance degradation of cyber-physical systems under stealthy integrity attacks. *IEEE Transactions on Automatic Control*, 61:2618–2624, 2016.

[25] Carlos Murguia and Justin Ruths. Characterization of a cusum model-based sensor attack detector. In *proceedings of the 55th IEEE Conference on Decision and Control (CDC)*, 2016.

[26] Carlos Murguia and Justin Ruths. Cusum and chi-squared attack detection of compromised sensors. In *proceedings of the IEEE Multi-Conference on Systems and Control (MSC)*, 2016.

[27] Carlos Murguia and Justin Ruths. On reachable sets of hidden cps sensor attacks. In *proceedings of the American Control Conference (ACC), 2018*, 2018.

[28] Carlos Murguia, Nathan van de Wouw, and Justin Ruths. Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools. In *proceedings of the IFAC World Congress*, 2016.

[29] F. Pasqualetti, F. Dorfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58:2715–2729, 2013.

[30] M. Ross. *Introduction to Probability Models, Ninth Edition*. Academic Press, Inc., Orlando, FL, USA, 2006.

[31] C. Scherer, P. Gahinet, and M. Chilali. Multiobjective output-feedback control via lmi optimization. *IEEE Transactions on Automatic Control*, 42:896–911, 1997.

[32] C. Scherer and S. Weiland. *Linear matrix inequalities in control*. Springer-Verlag, The Netherlands, 2000.

[33] J. Michael Steele. *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge University Press, New York, NY, USA, 2004.

[34] Zhanghan Tang, Margreta Kuijper, Michelle S. Chong, Iven Mareels, and Christopher Leckie. Linear system security—detection and correction of adversarial sensor attacks in the noise-free case. *Automatica*, 101:53 – 59, 2019.

[35] A. Teixeira, I. Shames, H. Sandberg, and H. Johansson. A secure control framework for resource-limited adversaries. *Automatica*, 51:135 – 148, 2015.

[36] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson. Secure control systems: A quantitative risk management approach. *IEEE Control Systems Magazine*, 35:24–45, 2015.

[37] A. van der Schaft. *L2-Gain and Passivity Techniques in Nonlinear Control*. Springer, Berlin, 1999.

A Monitor Design

We use Corollary 1 to obtain outer time-varying ellipsoidal approximations of the reachable set of the estimation error (13) driven by v_k and η_k in the attack-free case ($\delta_k = \mathbf{0}$). Once we have this ellipsoid, we project it onto the residual hyperplane to get the ellipsoid $r_k^T \Pi r_k = 1$ of the monitor. Denote by $\psi^e(k, e_1, \eta(\cdot), v(\cdot))$ the solution of (13) at time k given the initial estimation error e_1 and the infinite disturbance sequences $\eta(\cdot) := \{\eta_1, \eta_2, \ldots\}$ and $v(\cdot) := \{v_1, v_2, \ldots\}$. The reachable set we seek to quantify is given by

$$\mathcal{R}_k^e := \left\{ e \in \mathbb{R}^n \;\middle|\; \begin{array}{l} e = \psi^e(k, e_1, \eta(\cdot), v(\cdot));\ e_1 \in \mathbb{R}^n, \\ v_k^T v_k \leq \bar{v},\ \eta_k^T \eta_k \leq \bar{\eta},\ \forall\, k \in \mathbb{N}. \end{array} \right\}. \tag{A.1}$$

Lemma 9 *Consider the estimation error dynamics* (13) *with matrices* (A^p, C^p, E, F, L)*, the perturbation bounds* $\bar{v}, \bar{\eta} \in \mathbb{R}_{>0}$*, and assume no attacks to the system, i.e.,* $\delta_k = \mathbf{0}$*. For a given* $a \in (0, 1)$*, if there exist constants* $a_1 = a_1^*, \ldots, a_N = a_N^*$ *and matrix* $\mathcal{P} = \mathcal{P}^*$ *solution of* (8) *with* $A = (A^p - LC^p)$*,* $N = 2$*,* $B^1 = -LF$*,* $B^2 = E$*,* $W_1 = (1/\bar{\eta})I_m$*,* $W_2 = (1/\bar{v})I_n$*,* $p_1 = m$*, and* $p_2 = n$*; then,* $\mathcal{R}_k^e \subseteq \mathcal{E}_k^e := \{e \in \mathbb{R}^n | e^T \mathcal{P}^e e \leq \alpha_k^e\}$*, with* $\mathcal{P}^e = \mathcal{P}^*$ *and* $\alpha_k^e := a^{k-1} e_1^T \mathcal{P}^e e_1 + \big((2-a)(1-a^{k-1})\big)/(1-a)$ *, and the ellipsoid* $\mathcal{E}_k^e$ *has minimum volume in the sense of Corollary 1.*

Proof: The result follows Corollary 1. ■

By Lemma 9, the trajectories of the estimation error dynamics are contained in the time-varying ellipsoid $e^T \mathcal{P}^e e = \alpha_k^e$. Having this ellipsoid, we look for the matrix Π of the monitor leading to the minimum-volume ellipsoid $r^T \Pi r = 1$ satisfying, for $k \geq k^*$ and some $k^* \in \mathbb{N}$, $r_k^T \Pi r_k = (C^p e_k + \eta_k)^T \Pi (C^p e_k + \eta_k) \leq 1$ for $e_k \in \mathcal{E}_k^e$ and η_k such that $\eta_k^T \eta_k \leq \bar{\eta}$.

Proposition 3 *Consider the function* α_k^e *defined in Lemma 9 and define* $\alpha_\infty^e := \lim_{k\to\infty} \alpha_k^e = (2-a)/(1-a)$*. For every* $\epsilon \in \mathbb{R}_{>0}$*, there exists* $k^*(a, \epsilon, e_1, \mathcal{P}^e) \in \mathbb{N}$ *such that* $\alpha_k^e \leq \alpha_\infty^e + \epsilon$ *for all* $k \geq k^*(a, \epsilon, e_1, \mathcal{P}^e)$*.*

Proof: The function α_k^e can be written in terms of the

constant α^e_∞ as $\alpha^e_k = a^{k-1}e_1^T\mathcal{P}^*e_1 + (1-a^{k-1})\alpha^e_\infty$. Moreover, $\alpha^e_k \leq \alpha^e_\infty + \epsilon \Leftrightarrow \alpha^e_k - \alpha^e_\infty \leq \epsilon$ and $\alpha^e_k - \alpha^e_\infty = a^{k-1}\big(e_1^T\mathcal{P}^*e_1 - \alpha^e_\infty\big)$. Because $a<1$, inequality $a^{k-1}\big(e_1^T\mathcal{P}^*e_1 - \alpha^e_\infty\big) \leq \epsilon$, can always be satisfied for any $\epsilon \in \mathbb{R}_{>0}$ and sufficiently large k. ■

Remark 10 *By Proposition 3, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $k^* \in \mathbb{N}$ such that $\alpha^e_k \leq \alpha^e_\infty + \epsilon$ for all $k \geq k^*$. The least k^* satisfying $\alpha^e_{k^*} \leq \alpha^e_\infty + \epsilon$ is given by $k^*(a,\epsilon,e_1,\mathcal{P}^e) = \min\{k \in \mathbb{N} | a^{k-1}\big(e_1^T\mathcal{P}^e e_1 - \alpha^e_\infty\big) \leq \epsilon\}$ (see the proof of Proposition 3 above). Notice that $\alpha^e_k \leq \alpha^e_\infty + \epsilon$ for $k \geq k^*$ implies $\mathcal{E}^e_k \subseteq \mathcal{E}^e_\epsilon$, where $\mathcal{E}^e_\epsilon := \{e \in \mathbb{R}^n | e^T\mathcal{P}^e e \leq \alpha^e_\infty + \epsilon\}$, for all $k \geq k^*$. It follows that, for any $\epsilon > 0$, the estimation error e_k is contained in ellipsoid $e^T\mathcal{P}^e e = \alpha^e_\infty + \epsilon$ for $k \geq k^*$, i.e., $\mathcal{R}^e_k \subseteq \mathcal{E}^e_\epsilon \; \forall \; k \geq k^*$. Therefore, for a fixed ϵ (and corresponding k^*), the problem of finding Π of the monitor amounts to finding Π such that $(C^p e_k + \eta_k)^T\Pi(C^p e_k + \eta_k) \leq 1$ for all e_k and η_k satisfying $e_k^T\mathcal{P}^e_{\eta,v}e_k \leq \alpha^e_\infty + \epsilon$ and $\eta_k^T\eta_k \leq \bar{\eta}$. This can be posed as a convex optimization problem using the $\mathcal{S}$-procedure.*

Proposition 4 *Let the conditions of Lemma 9 be satisfied and consider the corresponding matrix $\mathcal{P}^e \in \mathbb{R}^{n\times n}$, the function α^e_k, the constant $\alpha^e_\infty = \lim_{k\to\infty}\alpha^e_k = (2-a)/(1-a)$, and some $\epsilon \in \mathbb{R}_{>0}$. If there exist $\tau_1, \tau_2 \in \mathbb{R}$ and $\Pi \in \mathbb{R}^{m\times m}$ solution of the following convex optimization:*

$$\begin{cases} \min\limits_{\Pi,\tau_1,\tau_2} \; -\log\det[\Pi], \\ s.t.\ \Pi \geq \mathbf{0},\ \tau_1 \geq 0,\ \tau_2 \geq 0,\ \textit{and} \\ \begin{bmatrix} f_1 & -(C^p)^T\Pi & \mathbf{0} \\ -\Pi C^p & \tau_2 I_m - \Pi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & f_2 \end{bmatrix} \geq \mathbf{0}, \\ f_1 = \tau_1\mathcal{P}^e - (C^p)^T\Pi C^p, \\ f_2 = 1 - \tau_1(\alpha^e_\infty + \epsilon) - \tau_2\bar{\eta}; \end{cases} \tag{A.2}$$

then, for $\delta_k = \mathbf{0}$ and $k \geq k^*(a,\epsilon,e_1,\mathcal{P}^e) = \min\{k \in \mathbb{N} | \alpha^e_k - \alpha^e_\infty \leq \epsilon\}$, the monitor inequality $r_k^T\Pi r_k \leq 1$ is satisfied for all e_k and η_k satisfying $e_k^T\mathcal{P}^e e_k \leq \alpha^e_\infty + \epsilon$ and $\eta_k^T\eta_k \leq \bar{\eta}$.

Proof: By Lemma 9, Proposition 3, and Remark 10, for any $\epsilon \in \mathbb{R}_{>0}$ and corresponding k^* satisfying $\alpha^e_{k^*} - \alpha^e_\infty \leq \epsilon$, the trajectories of estimation error dynamics (13) satisfy $e_k^T\mathcal{P}^e e_k \leq \alpha^e_\infty + \epsilon$ for all $k \geq k^*$. By the $\mathcal{S}$-procedure [4], if there exist $\tau_1, \tau_2 \in \mathbb{R}_{\geq 0}$ satisfying

$$\begin{aligned} &(C^p e_k + \eta_k)^T\Pi(C^p e_k + \eta_k) - 1 \\ &- \tau_1(e_k^T\mathcal{P}^e e_k - \alpha^e_\infty - \epsilon) - \tau_2(\eta_k^T\eta_k - \bar{\eta}) \leq 0, \end{aligned} \tag{A.3}$$

then, $(C^p e_k + \eta_k)^T\Pi(C^p e_k + \eta_k) \leq 1$ is satisfied for all e_k and η_k satisfying $e_k^T\mathcal{P}^e e_k \leq \alpha^e_\infty + \epsilon$ and $\eta_k^T\eta_k \leq \bar{\eta}$. Inequality (A.3) can be written as

$$v_k^T \underbrace{\begin{bmatrix} f_1 & -(C^p)^T\Pi & \mathbf{0} \\ -\Pi C^p & \tau_2 I_m - \Pi & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & f_2 \end{bmatrix}}_{Q} v_k \geq 0,$$

with $v_k := \big(e_k^T, \eta_k^T, 1\big)^T$. The above inequality is satisfied if and only if Q is positive semidefinite. Therefore, for $k \geq k^*$, $r_k^T\Pi r_k \leq 1$ for any Π solution of (A.2). Again, to ensure that the ellipsoidal bound is as tight as possible, we minimize $\log\det[\Pi^{-1}]$ as this objective shares the same minimizer with $(\det[\Pi])^{-1/2}$ and because for a positive definite Π it is convex [4]. ■

Remark 11 *Using Proposition 4, we can design monitors for every $\epsilon \in \mathbb{R}_{>0}$. If we want tight monitors, we need small ϵ because $\epsilon \approx 0$ yields $\mathcal{E}^e_k \subseteq \mathcal{E}^e_\epsilon \approx \mathcal{E}^e_\infty$ for $k \geq k^*$. That is, the contribution of initial conditions to the outer bound $\mathcal{E}^e_\epsilon$ on $\mathcal{E}^e_k$ used in Proposition 4 (see Remark 10) to compute the monitor matrix Π has decreased to a small value and mainly the effect of the perturbations η_k and v_k is taken into account when designing the monitor matrix Π. However, depending on the initial conditions, too small ϵ might result on very large k^*. The values of ϵ and k^* are related through the expression $k^* = \min\{k \in \mathbb{N} | \alpha^e_k - \alpha^e_\infty = a^{k-1}\big(e_1^T\mathcal{P}^*e_1 - \alpha^e_\infty\big) \leq \epsilon\}$ introduced in Proposition 4. Note that, for $e_1^T\mathcal{P}^*e_1 \leq \alpha^e_\infty$, $k^* = 1$ for any $\epsilon \in \mathbb{R}_{>0}$, i.e, ϵ can be selected arbitrarily small. On the other hand, $e_1^T\mathcal{P}^*e_1 > \alpha^e_\infty$ implies that $k^* \to \infty$ as $\epsilon \to 0$. That is, in this case, there is a trade-off between conservative monitors and convergence time when selecting ϵ.*

A.1 Proof of Lemma 5

Assume that the conditions of Lemma 5 are satisfied for some $a \in (0,1)$, $\epsilon \in \mathbb{R}_{>0}$, $a_1, a_2 \in \mathbb{R}$, and matrices (S, G, R). Because $L = S^{-1}R$ and $\Pi = G$, then $R = SL$, $G = \Pi$, and the matrix inequalities in (48) take the form:

$$S > \mathbf{0}, \begin{bmatrix} aS & (A^p - LC^p)^T S & \mathbf{0} & \mathbf{0} \\ S(A^p - LC^p) & S & -SLF & SE \\ \mathbf{0} & -(LF)^T S & \frac{1-a_1}{\bar{\eta}} I_m & \mathbf{0} \\ \mathbf{0} & E^T S & \mathbf{0} & \frac{1-a_2}{\bar{v}} I_n \end{bmatrix} \geq \mathbf{0}, \tag{A.4}$$

$$\Pi > \mathbf{0}, \begin{bmatrix} \frac{1}{\alpha^e_\infty + \epsilon + \bar{\eta}} S - (C^p)^T\Pi C^p & -(C^p)^T\Pi \\ -\Pi C^p & \frac{1}{\alpha^e_\infty + \epsilon + \bar{\eta}} I_m - \Pi \end{bmatrix} \geq \mathbf{0}. \tag{A.5}$$

The inequalities in (A.4) are of the form (6) in Proposition 1 with $\mathcal{P} = S$, $A = (A^p - LC^p)$, $N = 2$, $B^1 = -LF$, $B^2 = E$, $W_1 = (1/\bar{\eta})I_m$, $W_2 = (1/\bar{v})I_n$, $p_1 = m$, and $p_2 = n$. Hence, because $a, a_1, a_2 \in (0,1)$ and $a_1 + a_2 \geq a$, by Proposition 1, $e_k^T S e_k \leq \alpha^e_k$ for all $k \in \mathbb{N}$, e_k solution of (15) with $\delta_k = \mathbf{0}$, $\alpha^e_k = a^{k-1}e_1^T S e_1 + \alpha^e_\infty(1 - a^{k-1})$, and $\alpha^e_\infty = (2-a)/(1-a)$. Note that, for every $\epsilon > 0$, we have $\alpha^e_k \leq \alpha^e_\infty + \epsilon \Leftrightarrow \alpha^e_k - \alpha^e_\infty = a^{k-1}\big(e_1^T S e_1 - \alpha^e_\infty\big) \leq \epsilon$, and thus, because $a \in (0,1)$, $\alpha^e_k \leq \alpha^e_\infty + \epsilon$ for all $k \geq k^*(a,\epsilon,e_1,S) = \min\{k \in \mathbb{N} | a^{k-1}\big(e_1^T S e_1 - \alpha^e_\infty\big) \leq \epsilon\}$. Inequality $\alpha^e_k \leq \alpha^e_\infty + \epsilon$ for $k \geq k^*$ implies $e_k^T S e_k \leq \alpha^e_\infty + \epsilon$ for $k \geq k^*$, i.e, for any $\epsilon > 0$, the estimation error e_k satisfies $e_k^T S e_k \leq \alpha^e_\infty + \epsilon$ for all $k \geq k^*$. Moreover, because $\eta_k^T\eta_k \leq \bar{\eta}$ for $k \in \mathbb{N}$, it is easy to verify

that $w_k^T Q_1 w_k \leq q$ for $k \geq k^*$, where $w_k := (e_k^T, \eta_k^T)^T$, $Q_1 := \mathrm{diag}[S, I_m] > \mathbf{0}$, and $q := \alpha_\infty^e + \epsilon + \bar{\eta} \in \mathbb{R}_{>0}$. Since $r_k = C^p e_k + \eta_k$, the monitor inequality, $r_k^T \Pi r_k \leq 1$, can be written in terms of w_k as $w_k^T Q_2 w_k \leq 1$, where

$$Q_2 := \begin{bmatrix} (C^p)^T \Pi C^p & (C^p)^T \Pi \\ \Pi C^p & \Pi \end{bmatrix}.$$

Note that $w_k^T Q_1 w_k \leq q \Leftrightarrow w_k^T \left(\frac{1}{q} Q_1\right) w_k \leq 1$, because $q \in \mathbb{R}_{>0}$ and $Q_1 > \mathbf{0}$, and thus, if $w_k^T Q_2 w_k \leq w_k^T \left(\frac{1}{q} Q_1\right) w_k$, then $w_k^T Q_2 w_k \leq 1$ for $k \geq k^*$ (because $w_k^T Q_1 w_k \leq q$ only for $k \geq k^*$). Inequality $w_k^T Q_2 w_k \leq w_k^T \left(\frac{1}{q} Q_1\right) w_k$ is satisfied for any $w_k \in \mathbb{R}^{n+m}$ if and only if $\frac{1}{q} Q_1 - Q_2 \geq \mathbf{0}$. The latter inequality equals the right-hand side inequality in (A.5) and it is satisfied by assumption. Therefore, $w_k^T Q_2 w_k = r_k^T \Pi r_k \leq 1$ for $k \geq k^*$, $\Pi = G$, $L = S^{-1} R$, and $(a, a_1, a_2, \epsilon, S, G, R)$ satisfying (48). ■

A.2 Proof of Lemma 8

Let ν be such that $\tilde{\mathbf{X}}(\nu) > \mathbf{0}$ and $\mathbf{S}(\nu) \geq \mathbf{0}$. Because $\tilde{\mathbf{X}}(\nu) > \mathbf{0}$, by the Schur complement, $Y > 0$ and $X - Y^{-1} > 0$. Since $YX + VU^T = I$ (see (34)), then $VU^T = I - YX < \mathbf{0}$, i.e., the matrix VU^T is invertible. Hence, it is always possible to factorize $I - YX$ as $VU^T = I - YX$ with square and nonsingular U and V. Invertible U and V implies that $\mathcal{Y}$ and the matrix $\mathcal{T}_3 := \mathrm{diag}[\mathcal{Y}, \mathcal{Y}, I, I]$ are invertible. It follows that the transformations $\mathcal{X} \to \mathcal{Y}^T \mathcal{X} \mathcal{Y} = \tilde{\mathbf{X}}(\nu)$ and $\mathcal{S} \to \mathcal{T}_3^T \mathcal{S} \mathcal{T}_3 = \mathbf{S}(\nu)$ are congruent. Therefore, $\tilde{\mathbf{X}}(\nu) > \mathbf{0}$ and $\mathbf{S}(\nu) \geq \mathbf{0}$ imply $\mathcal{X} > 0$ and $\mathcal{S} \geq \mathbf{0}$ because $\tilde{\mathbf{X}}(\nu)$ and $\mathbf{S}(\nu)$ have the same signature as $\mathcal{X}$ and $\mathcal{S}$, respectively. Because $\mathbf{X}(\nu) > \mathbf{0}$, the matrices U and V are nonsingular. The latter implies that the change of variables in (40a) and $\mathcal{Y}$ are invertible and lead to unique $(\mathcal{X}, A^c, B^c, C^c, D^c)$ by inverting (40a) and $\tilde{\mathbf{X}}(\nu) = \mathcal{Y}^T \mathcal{X} \mathcal{Y}$ in (53), and, by Lemma 7, this (A^c, B^c, C^c, D^c) leads to $\sup_{d_k \in \mathcal{L}_2, d_k \neq \mathbf{0}} (\|s_k\|_2 / \|d_k\|_2) \leq \gamma$ for $\tilde{\zeta}_1 = \mathbf{0}$. ■

A.3 Projection of High Dimensional Ellipsoids onto Coordinate Hyperplanes

Lemma 10 *Consider the ellipsoid:*

$$\mathcal{E} := \left\{ x \in \mathbb{R}^n, y \in \mathbb{R}^m \;\middle|\; \begin{bmatrix} x \\ y \end{bmatrix}^T \underbrace{\begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix}}_{Q} \begin{bmatrix} x \\ y \end{bmatrix} = \alpha \right\},$$

for some positive definite matrix $Q \in \mathbb{R}^{(n+m)\times(n+m)}$ and constant $\alpha \in \mathbb{R}_{>0}$. The projection $\mathcal{E}'$ of $\mathcal{E}$ onto the x-hyperplane is given by the ellipsoid:

$$\mathcal{E}' := \left\{ x \in \mathbb{R}^n \,\middle|\, x^T [Q_1 - Q_2 Q_3^{-1} Q_2^T] x = \alpha \right\}.$$

Proof: The matrix Q is positive definite and thus $Q_1 \in \mathbb{R}^{n\times n}$ and $Q_3 \in \mathbb{R}^{m\times m}$ are nonsingular. It follows that Q can be factorized as:

$$\begin{bmatrix} Q_1 & Q_2 \\ Q_2^T & Q_3 \end{bmatrix} = \begin{bmatrix} I_n & \mathbf{0} \\ -Q_3^{-1} Q_2^T & I_m \end{bmatrix}^T \begin{bmatrix} Q_1 - Q_2 Q_3^{-1} Q_2^T & \mathbf{0} \\ \mathbf{0} & Q_3 \end{bmatrix} \times \begin{bmatrix} I_n & \mathbf{0} \\ -Q_3^{-1} Q_2^T & I_m \end{bmatrix}.$$

Introduce the change of coordinates:

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} := \begin{bmatrix} I_n & \mathbf{0} \\ -Q_3^{-1} Q_2^T & I_m \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{A.6}$$

In these coordinates, the ellipsoid $\mathcal{E}$ is given by

$$\mathcal{E} = \left\{ \begin{matrix} \bar{x} \in \mathbb{R}^n \\ \bar{y} \in \mathbb{R}^m \end{matrix} \;\middle|\; \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}^T \underbrace{\begin{bmatrix} Q_1 - Q_2 Q_3^{-1} Q_2^T & \mathbf{0} \\ \mathbf{0} & Q_3 \end{bmatrix}}_{\bar{Q}} \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \alpha \right\}.$$

The matrix $\bar{Q}$ is block diagonal; therefore, in the new coordinates, the *projection* of $\mathcal{E}$ onto $\bar{y} = \mathbf{0}$ (the $\bar{x}$-hyperplane) and the *intersection* of $\mathcal{E}$ with $\bar{y} = \mathbf{0}$ are *equal*. The intersection with $\bar{y} = \mathbf{0}$ (and thus the projection onto $\bar{y} = \mathbf{0}$) is simply given by $\mathcal{E}^{\bar{x}} := \{(\bar{x}, \bar{y}) \in \mathcal{E} | \bar{y} = \mathbf{0}\} = \{\bar{x} \in \mathbb{R}^n | \bar{x}^T [Q_1 - Q_2 Q_3^{-1} Q_2^T] \bar{x} = \alpha\}$. This $\mathcal{E}^{\bar{x}}$ provides an expression for all the points of $\mathcal{E}$ that lie on the $\bar{x}$-hyperplane. However, from (A.6), note that $\bar{x} = x$; therefore, $\mathcal{E}^{\bar{x}} = \mathcal{E}'$ and $\mathcal{E}'$ provides the locus for all the points of $\mathcal{E}$ that lie on the x-hyperplane. ■