# Safety monitoring under stealthy sensor injection attacks using reachable sets

# Safety monitoring under stealthy sensor injection attacks using reachable sets

**Cédric Escudero** * **Michelle S. Chong** ** **Paolo Massioni** *
**Eric Zamaï** *

* *Univ Lyon, INSA Lyon, Université Claude Bernard Lyon 1, Ecole Centrale de Lyon, CNRS, Ampère, UMR5005, 69621 Villeurbanne, France (e-mail: cedric.escudero@insa-lyon.fr, paolo.massioni@insa-lyon.fr, eric.zamai@insa-lyon.fr)*
** *Eindhoven University of Technology, (e-mail: m.s.t.chong@tue.nl)*

**Abstract:** Stealthy sensor injection attacks are serious threats for industrial plants as they can compromise the plant's integrity without being detected by traditional fault detectors. In this manuscript, we study the possibility of revealing the presence of such attacks by monitoring only the control input. This approach consists in computing an ellipsoidal bound of the input reachable set. When the control input does not belong to this set, this means that a stealthy sensor injection attack is driving the plant to critical states. The problem of finding this ellipsoidal bound is posed as a convex optimization problem (convex cost with Linear Matrix Inequalities constraints). Our monitoring approach is tested in simulation.

*Keywords:* Secure control, sensor attack, stealthy attack, deception attack, detection

## 1. INTRODUCTION

Information and communication technologies are progressively integrated in many industrial applications. In particular, the control of plants is targeted by this integration to improve the control performance and the awareness of their surrounding environment in making appropriate decisions. As a result, plants are now controlled by embedded computer devices over communication networks, commonly called Network Control Systems (NCSs) (Lee (2008)). Beyond the benefits made possible by information and communication technologies, new issues have emerged. A critical one is recently highlighted by academics as well as industrial stakeholders: the security issue (Cardenas et al. (2009)). NCSs are exposed to vulnerabilities (e.g. for remote control or maintenance of the embedded computer system via the internet or cellular networks), and can be exploited by adversaries to launch cyber attacks (Firoozjaei et al. (2022)). In particular, *deception* attacks have attracted the attention of the control engineering community. These attacks aim at degrading the control performance by tampering with system's signals (sensing and control). This includes spoofing, false-data injection, and replay attacks (Teixeira et al. (2012)) that can have severe consequences on the plant.

The secure control literature (Chong et al. (2019); Sandberg et al. (2022)) has rapidly grown to come up with different methods to detect deception attacks. Some works have proposed quantifying the impact of deception attacks on the plant by means of set-theoretic methods (Milošević et al. (2020); Liu et al. (2021); Zhang et al. (2022)). Set-theoretic methods involve computing sets to encapsulate the reachable states of dynamical systems; this is the so-called reachability analysis. Other works, based on reachability analysis, have proposed methods to mitigate the impact of stealthy attacks. Stealthy attacks are attacks that aim to damage the plant while avoiding dectection (e.g. fault detectors). Murguia et al. (2020) and Li et al. (2022) propose synthesis methods for the control block (i.e. the controller and the fault detector) to minimize the impact of stealthy attacks. Besides redesigning the control block, some works design monitoring approaches to reveal the presence of attacks. Among those works, Azzam et al. (2022) proposes a predictive approach to check the safety of the plant in the presence of potential stealthy attacks. Yang et al. (2021) uses an observer-based estimator using a bank of unknown input observers to estimate the plant states and detect attacks. Hu et al. (2019) reveals the presence of stealthy attacks by analyzing the residual distribution.

In this manuscript, we focus on *stealthy sensor injection attacks*, a class of deception attacks that injects malicious sensing signals into the network while escaping detection by the fault detector. We assume an adversary is capable of manipulating sensor measurements by compromising the communication network between the plant and the controller. The closest work relative to such attacks is in Murguia et al. (2020) where the authors model the attack and synthesize the control block to mitigate the impact of attacks on the plant. Here instead, we want to reveal the presence of stealthy sensor injection attacks by monitoring the control signals only. Indeed, sensor injection attacks will, at some time, affect the control signals to drive the plant into critical states, which will damage the plant integrity. Moreover, compared to estimator-based approaches, our approach will only require one signal to monitor (i.e., the control signal), hence reducing the vulnerabilities of the monitoring approach. To do so, we

propose a set-theoretic method based on reachable sets. It consists first in finding the state reachable set when the system is subject to stealthy sensor attacks and the plant state are constrained by a given safe set. The safe set is the set of plant states where its safe and proper operation is guaranteed. The resulting state reachable set encompasses all the trajectories of the plant states under stealthy attacks that are safe. Then, we search for an input reachable set that guarantees the plant states belong to the computed state reachable set for all stealthy attacks. Hence, the monitoring approach consists in verifying whether the control signal does not belong to this input reachable set, which we can then verify that the control system is under stealthy attacks due to our computationally efficient outer-approximation of the input reachable set. The computation of tight bounds on reachable sets is challenging in general (see Kurzhanski and Varaiya (2000)), especially when online computation is paramount, as is the case for critical control systems.

The rest of this manuscript is organized as follows. Section 2 formulates the research problem we want to address. Section 3 provides the necessary background about computing the ellipsoidal bound of reachable sets. Our approach of monitoring the input signals is presented in Section 4. Lastly in Section 5, we apply our results on a three-tank system to illustrate the performance of our monitoring approach.

***Notation:*** The symbol $\mathbb{R}$ stands for the real numbers, $\mathbb{R}^{n\times m}$ is the set of real $n \times m$ matrices, and $\mathbb{R}_{>0}$ ($\mathbb{R}_{\geq 0}$) denotes the set of positive (non-negative) real numbers. Matrix $A^\top$ indicates the transpose of matrix $A$ and $\text{diag}(a_1, ..., a_n)$ corresponds to a diagonal matrix with diagonal elements $a_1, ..., a_n$. The identity matrix of dimension $n$ is denoted by $I_n$, and $\mathbf{0}$ is a matrix of only zeros of appropriate dimensions. The notation $A \succeq 0$ (resp. $A \preceq 0$) indicates that the matrix $A$ is positive (resp. negative) semidefinite, i.e., all the eigenvalues of the symmetric matrix $A$ are positive (resp. negative) or equal to zero, whereas the notation $A \succ 0$ (resp. $A \prec 0$) indicates the positive (resp. negative) definiteness, i.e., all the eigenvalues are strictly positive (resp. negative). The notation $\mathcal{E}_\varphi(\Phi, \bar{\phi})$ stands for an ellipsoidal set of dimension $n_\phi$ with shape matrix $\Phi \in \mathbb{R}^{n_\varphi \times n_\varphi}$, $\Phi = \Phi^\top \succ 0$ and centered at $\bar{\phi}$, i.e., $\mathcal{E}_\varphi(\Phi, \bar{\phi}) := \{\varphi \in \mathbb{R}^n \,|\, (\varphi - \bar{\phi})^\top \Phi (\varphi - \bar{\phi}) \leqslant 1\}$; if no center $\bar{\phi}$ is specified in the ellipsoid notation this means that the ellipsoid is centred at 0, i.e. $\mathcal{E}_\varphi(\Phi) := \{\varphi \in \mathbb{R}^n \,|\, \varphi^\top \Phi \varphi \leqslant 1\}$.

## 2. PROBLEM FORMULATION

In this section, we introduce the class of systems and attacks under study. Then, we introduce at a high-level our proposed solution to detect those attacks based on input monitoring, and formulate the research problem addressed in this manuscript.

### *2.1 System dynamics*

We consider a system $\Sigma_p$ with the following dynamics

$$\begin{aligned} \dot{x}_p(t) &= A_p x_p(t) + B_p u(t) \\ y(t) &= C_p x_p(t) + D_p u(t), \end{aligned} \tag{1}$$

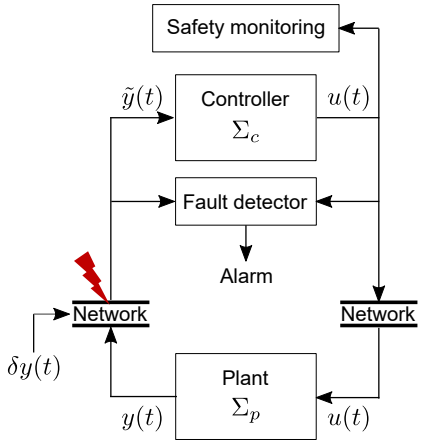


Fig. 1. Control system in the presence of sensor corruption.

with time $t \in \mathbb{R}_{>0}$, plant state $x_p \in \mathbb{R}^{n_p}$, sensor measurement $y \in \mathbb{R}^l$, control input $u \in \mathbb{R}^m$, and system matrices $A_p$, $B_p$, $C_p$, $D_p$ with appropriate dimensions.

The system description in (1) representing the plant is part of a closed-loop as illustrated in Figure 1. The system receives the control action $u$ and transmits the sensor measurement $y$ through a public/unsecured communication network.

### *2.2 Adversarial capabilities*

In this manuscript, we consider an adversary who can inject sensor data from the unsecured network. That is, it can add some signals $\delta y(t)$ to the true sensor measurement $y(t)$, i.e. $\delta y(t)$ models the corruption of sensor measurements. The adversary can compromise up to $s$ sensor measurements $y(t)$, with $s \in \{1, \ldots, l\}$. We also denote by $\Gamma$, an attacker's selection matrix to choose how the additive signals $\delta y(t)$ affects the true sensor measurements as proposed in Murguia et al. (2020). The received sensor measurement, i.e., the true sensor measurement with the additional signals controlled by the adversary, denoted $\tilde{y}(t)$ takes the following form:

$$\tilde{y}(t) = y(t) + \Gamma \delta y(t), \tag{2}$$

where $\delta y(t) \in \mathbb{R}^s$ denotes additive sensor measurement.

Notice that the attacker's selection matrix $\Gamma$ depends on the attacker and defender capabilities. If the defender secures some output channels, for instance with encryption, any data injection attack will not affect the received decrypted data. Similarly, the attacker wants to reduce its attack cost, then it might select only a few channels to perform its attack. The worst-case scenario for the defender is when $\Gamma = I_l$, which is when the attacker attacks all sensors.

### *2.3 Dynamic output feedback controller*

A controller $\Sigma_c$ receives the sensor measurement $\tilde{y}$ through the unsecured communication network and computes control actions $u$ which are transmitted to the plant.

The control input to the plant $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$, affected by the sensor attacks, is given by a controller $\Sigma_c$ with the following dynamics

$$\dot{x}_c(t) = A_c x_c(t) + B_c \tilde{y}(t)$$
$$u(t) = C_c x_c(t) + D_c \tilde{y}(t), \quad (3)$$

with controller state $x_c \in \mathbb{R}^{n_c}$.

We state all the assumptions about the plant and controller models below.

*Assumption 1.* (Modelling assumptions). The plant (1) and controller (3) models satisfy the following:

(i) The matrix $I_{l\times l} - D_p D_c$ is invertible such that the closed-loop system (1) and (3) is well-posed.
(ii) The plant (1) interconnected with the controller (3) is asymptotically stable. □

When $D_p = \mathbf{0}$, condition (i) in Assumption 1 holds.

### *2.4 Fault detector*

Traditionally, control systems are equipped with fault detectors to detect the occurrence of faults in the plant. Most model-based fault detectors consist in computing the residual, that is the difference between the sensor measurement and the estimated measurement. Here, to estimate the plant state $x_p$, we use the following state estimator:

$$\dot{\hat{x}}(t) = A_p \hat{x}(t) + B_p u(t) + L(\tilde{y}(t) - \hat{y}(t))$$
$$\hat{y}(t) = C_p \hat{x}(t) + D_p u(t), \quad (4)$$

with the estimated plant state $\hat{x} \in \mathbb{R}^{n_p}$, and gain matrix $L \in \mathbb{R}^{n_p \times l}$.

Let $e(t)$ denote the state estimation error defined as $e(t) := x_p(t) - \hat{x}(t)$, and $r(t)$ denote the residual defined as $r(t) := \tilde{y}(t) - \hat{y}(t)$. From the plant and controller models in (1), (3), the detector's estimation error and the residual evolve as follows:

$$\dot{e}(t) = (A_p - LC_p)e(t) - L\Gamma\delta y(t)$$
$$r(t) = C_p e(t) + \Gamma\delta y(t), \quad (5)$$

*Assumption 2.* (Modelling assumptions). The fault detector (5) model satisfies the following:

(i) The pair $(A_p, C_p)$ is observable such that there exists an $L$ of appropriate dimensions such that $(A_p - LC_p)$ is Hurwitz. □

Therefore, under Assumption 2, when there are no sensor attacks ($\delta_y(t) = 0$, for all $t \geq 0$), the detector's estimation error $e$ and residual $r$ will converge to the origin. Thereby, indicating that there is indeed no fault.

In fault detection, the objective is to decide between the two following hypotheses:

(1) $\mathcal{H}_0$: the system is normal (without faults/attacks)
(2) $\mathcal{H}_1$: the system is faulty

We consider here a fault detector that raises an alarm if $r(t)^\top \Pi r(t) > 1$ at some time $t$, with $\Pi \succ 0$. This means that we are under the normal mode when $r(t)^\top \Pi r(t) \leq 1$, $\forall\, t \in \mathbb{R}_{>0}$.

### *2.5 Stealthiness and attack definition*

In this manuscript, we consider an adversary who wants to remain stealthy with respect to the fault detector in (5). That is, it wants to launch *stealthy* sensor injection attacks that aim to avoid raising an alarm in the fault detector, i.e., being in the normal mode $\mathcal{H}_0$. Stealthy attacks are constrained attacks in the sense that the adversary must carefully choose the attack signal $\delta y(t)$ such that $r(t)^\top \Pi r(t) \leq 1$, $\forall t \geq 0$. Stealthy attacks often cannot lead to quick in time and high impact damage to the plant, but are long term attacks (e.g. causing the rapid aging of equipment). We define the stealthy set $\mathcal{E}_r(\Pi)$ as

$$r(t)^\top \Pi r(t) \leq 1 \Leftrightarrow r(t) \in \mathcal{E}_r(\Pi) \quad (6)$$

If (6) is satisfied for any attack signal $\delta y(t)$, then no alarms will be raised in the fault detector, that is sensor injection attacks are stealthy.

From the residual dynamics in (5), it yields that $\delta y(t) = \Gamma^+(r(t) - C_p e(t))$, where $\Gamma^+$ denotes the Moore-Penrose inverse of $\Gamma$ (see Murguia et al. (2020) for more details). We can now state our closed-loop system from plant and controller dynamics in (1), (3) together with state estimation error of the fault detector in (5) by taking the extended state $\zeta := [x_p^\top, x_c^\top, e^\top]^\top$, with $\zeta \in \mathbb{R}^n$ where $n = n_p + n_c + n_p$:

$$\dot{\zeta}(t) = A\zeta(t) + Br(t) \quad (7)$$
$$u(t) = E\zeta(t) + Fr(t) \quad (8)$$

with $A$, $B$, $E$, $F$ defined in (9) and $\Lambda := (I_l - D_p D_c)^{-1}$.

The setup is illustrated in Figure 1.

We denote the solution to the closed-loop system with initial condition $\zeta(t_0)$ and sensor corruption $\delta y$ by $\zeta(t; \zeta(t_0), \delta y)$. In the absence of the measurement corruption $\delta y$, i.e., $\delta y(t) = 0$ for all $t \geq t_0$, we call system (7) our nominal closed-loop system. Under our modelling assumptions (Assumption 1), the origin is an equilibrium point of our nominal closed-loop system (7).

After having described the system under study with the class of attacks the system might face, we now need a degradation metric to define the safety level under stealthy sensor injection attacks. To this end, we define the notion of safe sets (Escudero et al. (2022b)).

*Definition 1.* (Safe set $\mathcal{X}_s$). The safe set $\mathcal{X}_s \subseteq \mathbb{R}^{n_p}$ for system in (1) is the set of states $x_p \in \mathcal{X}_s$ where the safe and proper operation of the system is guaranteed.

Safe sets exclude, by Definition 1, critical states of the plant (1). A critical state is a state of the plant that, if reached, compromises the integrity of the system. For instance, the over-pressure in gas pipelines, or the null-distance/negative distance between two vehicles leading to collision are critical states. We can now formally define stealthy sensor injection attacks.

*Definition 2.* (Stealthy sensor injection attacks). Attacks that tamper with sensor measurement by injecting signals $\delta y(t)$ to true sensor measurements, $y(t)$, and aim to degrade the operation of the system dynamics in (1) by pushing trajectories outside the safe set $\mathcal{X}_s$ while keeping the residuals trajectories inside the stealthy set $\mathcal{E}_r$ for stealthiness.

Notice that stealthy sensor injection attacks defined in Definition 2 only exist if some critical states can be reached while the fault detector does not raise an alarm.

$$A = \begin{bmatrix} A_p + B_p D_c \Lambda C_p & B_p C_c + B_p D_c \Lambda D_p C_c & -B_p D_c \Gamma\Gamma^+ C_p - B_p D_c \Lambda D_p D_c \Gamma\Gamma^+ C_p \\ B_c \Lambda C_p & A_c + B_c \Lambda D_p C_c & -B_c \Gamma\Gamma^+ C_p - B_c \Lambda D_p D_c \Gamma\Gamma^+ C_p \\ \mathbf{0} & \mathbf{0} & A_p - L C_p + L\Gamma\Gamma^+ C_p \end{bmatrix}, \; F = \begin{bmatrix} (D_c\Gamma + D_c\Lambda D_p D_c \Gamma)\Gamma^+ \end{bmatrix},$$

$$B = \begin{bmatrix} B_p D_c \Gamma\Gamma^+ + B_p D_c \Lambda D_p D_c \Gamma\Gamma^+ \\ B_c \Gamma\Gamma^+ + B_c \Lambda D_p D_c \Gamma\Gamma^+ \\ -L\Gamma\Gamma^+ \end{bmatrix}, \qquad E = \begin{bmatrix} D_c \Lambda C_p & C_c + D_c \Lambda D_p C_c & -(D_c\Gamma + D_c\Lambda D_p D_c\Gamma)\Gamma^+ C_p \end{bmatrix}. \quad (9)$$

In the absence of measurement corruption $\delta y(t)$, we assume that the controller $\Sigma_c$ in (3) is designed such that the state of the plant (1) evolves within a known safe state set $\mathcal{X}_s$, which is compact.

*Assumption 3.* (Safe set $\mathcal{X}_s$). Given a safe set $\mathcal{X}_s \subset \mathbb{R}^{n_p}$ which is compact, the state of the plant (1) in the nominal closed-loop system (7) (with $\delta y(t) = 0$ for all $t \geq t_0$) satisfies $x_p(t, x_p(t_0)) \in \mathcal{X}_s \subset \mathbb{R}^{n_p}$, for all $t \geq t_0$ and all initial conditions $x_p(t_0) \in \mathbb{R}^{n_p}$. Further, the equilibrium of the nominal closed-loop system satisfies $0 \in \mathcal{X}_s$. □

### 2.6 Safety monitoring

We propose to detect the presence of stealthy sensor injection attacks $\delta y$ which has not been detected by the fault detector (stealthy), when we know the plant, the controller, the fault-detector models in (1), (3), (4), the stealthy set (6), and the safe set $\mathcal{X}_s$ defined in Definition 2, while only having access to the input $u(t)$ that might be affected by a corrupted sensor signal $\delta y$. This goal is called *safety monitoring* under stealthy sensor corruption.

The approach proposed in this manuscript is to find an input set $\mathcal{U}$, in which if the control input $u(t)$ does not belong to this set then this means that a stealthy sensor injection attack is driving the plant to critical states. We can now define our research problem as follows.

*Problem 1.* Given the closed-loop system in (7), the stealthy set $\mathcal{E}_r$ in (6), and the safe set $\mathcal{X}_s$ defined in Definition 1, find an input set $\mathcal{U}$ such that if the control input $u(t)$ does not belong to this set $\mathcal{U}$ at some time $t \geq t_0$, i.e., $u(t) \notin \mathcal{U}$, then the system trajectories are not contained in $\mathcal{X}_s$ at some $t \geq t_0$ for all stealthy sensor injection attacks defined in Definition 2.

## 3. PRELIMINARY TOOL

### 3.1 Stealthy state reachable set

First, we introduce the definition of the stealthy state reachable set of the closed-loop system (7). This set will be used later on to find an input set $\mathcal{U}$ to solve Problem 1.

*Definition 3.* (Stealthy state reachable set). The stealthy state reachable set $\mathcal{R}_\zeta(t)$ at time $t \in \mathbb{R}_{>0}$ from initial condition $\zeta(t_0) \in \mathbb{R}^n$ is the set of extended state $\zeta(t)$ that satisfy the extended differential equations (7), over all residuals $r(t)$ satisfying $r(t) \in \mathcal{E}_r(\Pi)$ to guarantee the attack stealthiness, i.e.,

$$\mathcal{R}_\zeta(t) := \left\{ \zeta(t) \;\middle|\; \begin{array}{l} \zeta(t_0) \in \mathbb{R}^n, \\ \zeta(t) \text{ satisfies (7)}, \\ \text{and } r(t) \in \mathcal{E}_r(\Pi). \end{array} \right\}. \quad (10)$$

Let $\mathcal{R}_\zeta(\infty)$ denote the asymptotic stealthy state reachable set, that is the ultimate bound on $\mathcal{R}_\zeta(t)$, i.e., $\mathcal{R}_\zeta(\infty) := \lim_{t\to\infty} \mathcal{R}_\zeta(t)$.

*Remark 1.* Because $r(t)$ is bounded ($r(t) \in \mathcal{E}_r(\Pi)$), then the asymptotic stealthy state reachable set $\mathcal{R}_\zeta(\infty)$ is compact if $A$ in (7) is Hurwitz.

### 3.2 Ellipsoidal bound on $\mathcal{R}_\zeta(\infty)$

Because the computation of the asymptotic stealthy state reachable set $\mathcal{R}_\zeta(\infty)$ is not tractable, we propose to rely on an outer ellipsoidal approximation $\mathcal{E}_\zeta(Q)$ of $\mathcal{R}_\zeta(\infty)$, i.e., $\mathcal{R}_\zeta(\infty) \subseteq \mathcal{E}_\zeta(Q)$. For the sake of clarity, we will refer to $\mathcal{E}_\zeta(Q)$ as an ellipsoidal bound on $\mathcal{R}_\zeta(\infty)$. To find this ellipsoidal bound, we propose searching for an invariant set for the dynamical system (7), defined as follows.

*Definition 4.* (Ellipsoidal bound $\mathcal{E}_\zeta(Q)$). For $Q \succ 0$, the ellipsoidal set $\mathcal{E}_\zeta(Q)$ is invariant for the dynamical system in (7), if for all initial states $\zeta(t_0) \in \mathcal{E}_\zeta(Q)$, and for all $r(t) \in \mathcal{E}_r(\Pi)$, the trajectories of $\zeta(t)$ in (7) satisfy $\zeta(t) \in \mathcal{E}_\zeta(Q)$, $\forall t \geq t_0$.

Escudero et al. (2022a) provides sufficient conditions for ellipsoidal sets $\mathcal{E}_\zeta(Q)$ to be invariant for a class of LTI systems as in (7) and for some states $\zeta(t)$ that are constrained by a given ellipsoidal set $\mathcal{E}_\varphi(\Phi, \bar{\phi})$, i.e., $\zeta(t) \in \mathcal{E}_\varphi(\Phi, \bar{\varphi})$, with $\Phi \succeq 0$. The method searches for a Lyapunov-like function $V(\zeta) = \zeta^\top Q \zeta$ using Linear Matrix Inequalities (LMIs) (Boyd et al. (1994)).

We now state the preliminary tool used to find invariant ellipsoidal sets for the closed-loop system in (7) with $r(t) \in \mathcal{E}_r(\Pi)$, $\zeta(t) \in \mathcal{E}_\varphi(\Phi, \bar{\varphi})$, $\forall t \geq t_0$.

*Lemma 1.* (Invariant Ellipsoidal Set). Consider the closed-loop system as in (7). If there exist matrix $Q \in \mathbb{R}^{n\times n}$ and constants $\alpha$, $\beta$, $\lambda \in \mathbb{R}_{\geq 0}$ satisfying the following inequalities:

$$-H - \alpha J - \beta K - \lambda L \succeq 0, \quad (11)$$

$$Q \succ 0, \quad (12)$$

with

$$H = \begin{bmatrix} A^\top Q + QA & \mathbf{0} & QB \\ * & \mathbf{0} & \mathbf{0} \\ * & * & \mathbf{0} \end{bmatrix}, \; J = \begin{bmatrix} Q & \mathbf{0} & \mathbf{0} \\ * & -1 & \mathbf{0} \\ * & * & \mathbf{0} \end{bmatrix},$$

$$K = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & 1 & \mathbf{0} \\ * & * & -\Pi \end{bmatrix}, \; L = \begin{bmatrix} -\Phi & \Phi\bar{\varphi} & \mathbf{0} \\ * & 1 - \bar{\varphi}^\top \Phi \bar{\varphi} & \mathbf{0} \\ * & * & \mathbf{0} \end{bmatrix};$$

then, $\zeta(t_0)^\top Q \zeta(t_0) \leqslant 1 \Rightarrow \zeta(t)^\top Q \zeta(t) \leqslant 1$, for all $t \geq t_0$, $r(t) \in \mathcal{E}_r(\Pi)$, and $\zeta(t) \in \mathcal{E}_\varphi(\Phi, \bar{\varphi})$.

*Proof 1.* Consider first (11); left and right multiply by $[\zeta(t)^\top, 1, r(t)^\top]^\top$, and consider $\alpha J$, $\beta K$ and $\lambda L$ as S-procedure terms by positive multipliers $\alpha$, $\beta$ and $\lambda$; this implies:

$$[\zeta^\top, 1, r^\top]\, H\, [\zeta^\top, 1, r^\top]^\top = \dot{V}(\zeta) \leqslant 0. \tag{13}$$

when

$$[\zeta^\top, 1, r^\top]\, J\, [\zeta^\top, 1, r^\top]^\top = V(\zeta) - 1 \geqslant 0 \Leftrightarrow V(\zeta) \geqslant 1. \tag{14}$$

$$[\zeta^\top, 1, r^\top]\, K\, [\zeta^\top, 1, r^\top]^\top \geqslant 0 \Leftrightarrow r(t) \in \mathcal{E}_r(\Pi). \tag{15}$$

$$[\zeta^\top, 1, r^\top]\, L\, [\zeta^\top, 1, r^\top]^\top \geqslant 0 \Leftrightarrow \zeta(t) \in \mathcal{E}_\varphi(\Phi, \bar{\varphi}). \tag{16}$$

This means that the value of $V(\zeta)$ can only increase under the stated constraints, i.e., $V(\zeta(t_0)) \leqslant 1 \Rightarrow V(\zeta(t)) \leqslant 1$ $\forall t \geq t_0$, which concludes the proof.

## 4. SOLUTION TO PROBLEM 1

In this section, we propose a mathematical framework, built around Lemma 1, to find an input set $\mathcal{U}$ solving Problem 1.

Similarly to the stealthy state reachable set, we define the stealthy input reachable set $\mathcal{R}_u(t)$.

*Definition 5.* (Stealthy input reachable set). The stealthy input reachable set $\mathcal{R}_u(t)$ at time $t \in \mathbb{R}_{>0}$ from initial condition $u(t_0) \in \mathbb{R}^m$ is the set of input $u(t)$ that satisfy the equations (8), over all $\zeta(t)$ satisfying the differential equations (7) and over all residuals $r(t)$ satisfying $r(t) \in \mathcal{E}_r(\Pi)$ to guarantee the attack's stealthiness, i.e.,

$$\mathcal{R}_u(t) := \left\{ u(t) \left| \begin{array}{l} u(t_0) \in \mathbb{R}^m, \\ u(t) \text{ satisfies (8)}, \\ \text{and } \zeta(t) \text{ satisfies (7)}, \\ \text{and } r(t) \in \mathcal{E}_r(\Pi). \end{array} \right. \right\}. \tag{17}$$

Let $\mathcal{R}_u(\infty)$ denote the asymptotic stealthy input reachable set, that is the ultimate bound on $\mathcal{R}_u(t)$, i.e., $\mathcal{R}_u(\infty) := \lim_{t\to\infty} \mathcal{R}_u(t)$.

*Remark 2.* Because $r(t)$ is bounded ($r(t) \in \mathcal{E}_r(\Pi)$) and $\zeta(t)$ is bounded ($\zeta(t) \in \mathcal{R}_\zeta(\infty)$) if condition (ii) in Assumption 1 holds, then the asymptotic stealthy input reachable set $\mathcal{R}_u(\infty)$ exists.

In this manuscript, we propose to rely on an outer ellipsoidal approximation $\mathcal{E}_u(R)$ of $\mathcal{R}_u(\infty)$, i.e. $\mathcal{R}_u(\infty) \subseteq \mathcal{E}_u(R)$. For the sake of clarity, we will refer $\mathcal{E}_u(R)$ as an ellipsoidal bound on $\mathcal{R}_u(\infty)$, with $R \succ 0$.

Before presenting the proposed approach, we model the safe set $\mathcal{X}_s$ defined in Definition 1 as an ellipsoid $\mathcal{E}_s(\Psi, \bar{\psi})$ written in terms of the extended state $\zeta$ and satisfying:

$$(\zeta - \bar{\psi})^\top \Psi (\zeta - \bar{\psi}) \leq 1 \tag{18}$$

with

$$\Psi = \begin{bmatrix} \Psi_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \bar{\psi} = \begin{bmatrix} \bar{\psi}_p \\ \mathbf{0} \end{bmatrix} \tag{19}$$

for some known positive semi-definite matrix $\Psi_p \in \mathbb{R}^{n_p \times n_p}$ and vector $\bar{\psi}_p \in \mathbb{R}^{n_p}$. Note that $\Psi_p$ is in general rank-deficient as it only constrains some of the plant states $x_p(t)$ – $\mathcal{E}_s(\Psi, \bar{\psi})$ can even coincide with $\mathbb{R}^{n\times n}$ by picking $\Psi_p = \mathbf{0}$.

### 4.1 Approach

To solve Problem 1 (see Section 2.6), we propose a two-step procedure: first we compute an ellipsoidal bound $\mathcal{E}_\zeta(Q)$ on the asymptotic stealthy state reachable set of the closed-loop system (7) for some states $\zeta(t)$ constrained to remain inside the safe set $\mathcal{E}_s(\Psi, \bar{\psi})$. This ellipsoidal bound describes where the state trajectories $\zeta(t)$ will remain at all times $t \geq t_0$ for some initial conditions $\zeta(t_0) \in \mathcal{E}_\zeta(Q)$ under the presence of stealthy sensor attacks while some parts of the states $\zeta$ are constrained by the safe set $\mathcal{E}_s$. Once this state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ is computed, we compute the input ellipsoidal bound $\mathcal{E}_u(R)$ on $\mathcal{R}_u(\infty)$ when the states $\zeta(t)$ belong to the ellipsoidal bound $\mathcal{E}_\zeta(Q)$ and the residuals $r(t)$ belong to the stealthy set $\mathcal{E}_r(\Pi)$.

If such an input ellipsoidal bound $\mathcal{E}_u(R)$ can be computed, we can use it to check whether the input signal $u(t)$ is in it. If $u(t)$ does not belong to this set, then it means that the safety of the plant will be violated at some time $t \geq t_0$. Indeed, we only have an outer-approximation of the asymptotic stealthy state reachable set $\mathcal{R}_\zeta(\infty)$ as its exact computation is not tractable. As a result, the corresponding input set $\mathcal{E}_u(R)$ is also an outer-approximation of the asymptotic input reachable set.

### 4.2 Main theorem

The main result of this manuscript is the computation of an input ellipsoidal bound $\mathcal{E}_u(R)$ on $\mathcal{R}_\zeta(\infty)$ of the closed-loop system (7) under the presence of stealthy sensor injection attacks $\delta y(t)$ when the state trajectories $\zeta(t)$ are bounded by the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ at all time $t \geq t_0$, which represents where the state $\zeta(t)$ will remain at all time $t \geq t_0$ when they are constrained by the safe set under the presence of any stealthy sensor injection attacks.

The problem we want to solve is formulated as follows. We want to find an input ellipsoidal bound $\mathcal{E}_u(R)$ such that $u(t)$ satisfy the equation (8) for all $\zeta(t)$, $r(t)$ satisfying $\zeta(t) \in \mathcal{E}_\zeta(Q)$ and $r(t) \in \mathcal{E}_r(\Pi)$ at all time $t \geq t_0$. This means that if the input signal $u(t)$ does not belong to $\mathcal{E}_u(R, \bar{u})$, thus the input signal affected by a stealthy sensor injection attack is violating at some time $t$ at least one constraint, i.e. the state trajectories $\zeta(t)$ are constrained to remain inside the safe set.

*Theorem 1.* (Input ellipsoidal bound $\mathcal{E}_u(R)$). Consider the closed-loop system as in (7), an invariant ellipsoidal set $\mathcal{E}_\zeta(Q)$ with $Q \succ 0$ for the system in (7), and the stealthy set $\mathcal{E}_r(\Pi)$. If there exist matrix $R \in \mathbb{R}^{m\times m}$ and constants $\gamma, \tau \in \mathbb{R}_{\geq 0}$ satisfying the following inequalities:

$$-W - \gamma Y - \tau Z \succeq 0, \tag{20}$$

$$R \succ 0 \tag{21}$$

with

$$W = \begin{bmatrix} E^\top R E^\top & \mathbf{0} & E^\top R F \\ * & -1 & \mathbf{0} \\ * & * & F^\top R F \end{bmatrix}, \tag{22}$$

$$Y = \begin{bmatrix} -Q & \mathbf{0} & \mathbf{0} \\ * & 1 & \mathbf{0} \\ * & * & \mathbf{0} \end{bmatrix}, \tag{23}$$

$$Z = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & 1 & \mathbf{0} \\ * & * & -\Pi \end{bmatrix} \tag{24}$$

then, $u(t_0)^\top R u(t_0) \leqslant 1 \Rightarrow \zeta(t)^\top Q \zeta(t) \leqslant 1, r(t)^\top \Pi r(t) \leqslant 1$, for all $t \geq t_0$

*Proof 2.* Consider (20); left and right multiply by $[\zeta(t)^\top, 1, r(t)^\top]^\top$, and consider $\gamma Y$ and $\tau Z$ as S-procedure terms

by positive multipliers $\gamma$, and $\tau$; this implies with the S-procedure:

$$-[\zeta^\top, 1, r^\top]\, W\, [\zeta^\top, 1, r^\top]^\top \geq 0 \Leftrightarrow u(t) \in \mathcal{E}_u(R)$$

when

$$[\zeta^\top, 1, r^\top]\, Y\, [\zeta^\top, 1, r^\top]^\top \geq 0 \Leftrightarrow \zeta(t) \in \mathcal{E}_\zeta(Q),$$

$$[\zeta^\top, 1, r^\top]\, Z\, [\zeta^\top, 1, r^\top]^\top \geq 0 \Leftrightarrow r(t) \in \mathcal{E}_r(\Pi)$$

This means that the input trajectories remain inside the input ellipsoidal bound $\mathcal{E}_u(R)$ for all (i) plant state trajectories inside the invariant ellipsoidal set $\mathcal{E}_\zeta(Q)$ and (ii) residual trajectories $r(t)$ inside the stealthy set $\mathcal{E}_r(\Pi)$.

### *4.3 Computation of the input ellipsoidal bound $\mathcal{E}_u(R)$*

To compute the input ellipsoidal bound $\mathcal{E}_u(R)$, first we have to compute the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ using Lemma 1. Due to the product of $\alpha$ with $J$, the matrix inequality (11) in Lemma 1 is not an LMI. To relax it, we will compute the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ for a fixed $\alpha$. Because we compute an outer-approximation of the stealthy state reachable set $\mathcal{R}_\zeta(\infty)$, we want to find the one with the lowest volume to have the best approximation. Since the volume of an ellipsoid $\mathcal{E}_\varphi(\Phi)$ is proportional to $(\det(\Phi)^{-1/2}$, we can minimize $(\log(\det(\Phi))^{-1}$ to minimize the volume of the ellipsoid, which is convex and allows to cast the problem as a convex optimization problem (Boyd et al. (1994)). By solving the convex optimization problem $\mathbf{OP}_1$ for a fixed $\alpha$ we find a state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ with the lowest volume for the fixed parameter.

**$\mathbf{OP}_1$: State ellipsoidal bound $\mathcal{E}_\zeta(Q)$**

$$\begin{aligned} \underset{Q,\beta,\lambda}{\text{minimize}} \quad & \text{-log(det}(Q)), \\ \text{subject to} \quad & (11), (12). \end{aligned}$$

After having found the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$, we can now compute the input ellipsoidal bound $\mathcal{E}_u(R)$ using Theorem 1 by solving the convex optimization problem $\mathbf{OP}_2$. Similarly, we want to find an input ellipsoidal bound with the lowest volume.

**$\mathbf{OP}_2$: Input ellipsoidal bound $\mathcal{E}_u(R)$**

$$\begin{aligned} \underset{R,\gamma,\tau}{\text{minimize}} \quad & \text{-log(det}(R)), \\ \text{subject to} \quad & (20), (21). \end{aligned}$$

The procedure to find an input ellipsoidal bound $\mathcal{E}_u(R)$ answering Problem 1 is summarized in Algorithm 1.

**Algorithm 1:** Compute input ellipsoidal bound $\mathcal{E}_u(R)$

**Result:** Input ellipsoidal bound $\mathcal{E}_u(R)$
Init: closed-loop system ($A$, $B$, $E$, $F$), stealthy set ($\mathcal{E}_r(\Pi)$), safe set ($\Psi$, $\bar{\psi}$), attacker's selection matrix ($\Gamma$);
1) Find the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ by solving $\mathbf{OP}_1$ for some $\alpha \geq 0$;
2) Find the input ellipsoidal bound $\mathcal{E}_u(R)$ by solving $\mathbf{OP}_2$;

## 5. SIMULATION EXAMPLE

We verify our main result in simulations on a three-tank system. After stating the models of the three-tank system, the controller and the fault-detector, we first describe the attack scenario against the system. Then, we apply our main results using Algorithm 1. Lastly, we show that the proposed approach enables the detection of some stealthy attacks and we highlight the limitations of using reachable sets for detection. We use the solver MOSEK with the YALMIP toolbox on Matlab to solve the optimization problems, and we use Simulink to inject attacks on the system and test the proposed approach.

### *5.1 System description*

Consider the three-tank system from Iqbal et al. (2007) modelled in (25) as an LTI system as in (1) with $x_p = [x_{p1}, x_{p2}, x_{p3}]^\top$ ($n_p = 3$) and $u = [u_1, u_2]^\top$ ($m = 2$) where $x_p$ is the tank's liquid level [cm], and $u$ is the supply flow rates [mL.s$^{-1}$]. The plant states $x_{p1}$ and $x_{p2}$ are measured and encompassed in the sensor measurement $y = [y_1, y_2]^T$ ($l = 2$).

$$A_p = 10^{-4} \times \begin{bmatrix} -1.36 & 0 & 0.72 \\ 0 & -2.29 & 15.30 \\ 1.36 & -1.11 & -16.02 \end{bmatrix}, B_p = \begin{bmatrix} 64.94 & 0 \\ 0 & 64.94 \\ 0 & 0 \end{bmatrix}$$

$$C_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, D_p = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{25}$$

An observer-based output feedback controller as in (3) is considered in closed-loop with controller matrices given in (26) with $n_c = 3$.

$$A_c = \begin{bmatrix} -0.33 & 0.09 & -122.34 \\ 0.10 & -0.33 & 114.69 \\ -0.01 & -0.16 & -0.002 \end{bmatrix}, B_c = 10^{-2} \times \begin{bmatrix} 2.00 & 0.13 \\ 0.19 & 3.80 \\ 1.49 & 15.49 \end{bmatrix}$$

$$C_c = 10^{-2} \times \begin{bmatrix} -0.47 & 0.14 & -188.41 \\ 0.16 & -0.45 & 176.62 \end{bmatrix}, D_c = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \tag{26}$$

A fault detector is implemented as in Fig. 1 with gain matrix $L$ given below in (27). The fault detector raises an alarm when $r(t) \notin \mathcal{E}_r(\Pi)$ with $\Pi = \text{diag}(1, 1)$.

$$L = 10^{-3} \times \begin{bmatrix} 20.01 & 1.31 \\ 1.92 & 38.02 \\ 14.93 & 154.94 \end{bmatrix} \tag{27}$$

Notice that the gain matrix $L$ is chosen intentionally here to admit a large class of stealthy attack signals.

The three-tank system operates safely when the states remains inside the safe set $\mathcal{E}_s(\Psi, \bar{\psi})$ defined for $\Psi_p = \text{diag}(10^{-4}, 10, 10^{-4})$, and $\bar{\psi}_p = \mathbf{0}$. This means that the plant state $x_{p2}$ is constrained whereas $x_{p1}$ and $x_{p3}$ as the corresponding terms in $\Psi_p$ tend to zero.

### *5.2 Attack scenario*

We consider here the worst-case scenario where an adversary is capable of compromising both output measurements $y_1(t)$ and $y_2(t)$ ($s = 2$) by injecting two attack signals $\delta y_1$ and $\delta y_2$ added to the measurements, i.e., $\Gamma = I_2$.

### *5.3 Safety monitoring*

Here we want to find an input set $\mathcal{U}$ as a solution to Problem 1 using Algorithm 1. First, we compute the state

ellipsoidal bound for $\alpha = 30$. The projection of the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ onto the $x_p$-hyperplane is drawn on the left-hand side in Figure 2 (green fill) together with the safe set $\mathcal{E}_s$ (blue fill). As we can expect, the plant state $x_{p2}$ is constrained whereas $x_{p1}$ and $x_{p3}$ are not (they can reach almost the entire state space). Then, we compute the input ellipsoidal bound $\mathcal{E}_u(R)$ where the resulting $R$ is given in (28). The input ellipsoidal bound is drawn and zoomed in on the right-hand side in Figure 2 (red fill).

$$R = 10^5 \times \begin{bmatrix} 3.99 & 4.26 \\ 4.26 & 4.55 \end{bmatrix} \tag{28}$$

After having found the input ellipsoidal bound $\mathcal{E}_u(R)$, we test the proposed detection scheme based on the input ellipsoidal bound to verify if we can detect stealthy sensor injection attacks by checking whether the control input $u(t)$ belongs to the input ellipsoidal bound or not. Figure 3 shows the simulation result with, from the top to the bottom on the left-hand side: the plant states $x_p(t)$, a safety check (takes 1 if the plant states do not belong to the safe set; otherwise 0), the control input $u(t)$, and the detection result (returns 1 if the control input do not belong to the input ellipsoidal bound $\mathcal{E}_u(R)$; otherwise 0). In Figure 3, from the top to the bottom on the right-hand side, the attack signals $\delta y(t)$, the residuals $r(t)$, and a stealthiness check (returns 1 if the residuals belong to the stealthy set; otherwise 0) are drawn. Three attacks have been launched on the closed-loop system : first attack starts at 10 s and ends at 110 s, second attack starts at 175 s and ends at 275 s, and last attack starts at 340 s and ends at 440 s. As we can observe, the attack signals $\delta y$ are carefully choosen to keep the fault detector below its detection threshold, i.e. the residualts $r(t) \in \mathcal{E}_s(\Pi)$ (see the stealthiness check in Figure 3). Each attack violates the safety of the plant at some time $t$. For each attack, the safety is first violated respectively at 42.3 s, 188.1 s, and 388.5 s. As we can pinpoint, the attack signals $\delta y(t)$ affect the control input $u(t)$ to compromise the integrity of the plant. For each attack, the detection occurs at 59.8 s, 218.4 s, and no detection is triggered for the last attack.

First, we can see that the proposed approach succeeded in revealing the presence of stealthy sensor injection attacks by monitoring only the control input $u(t)$ for the first two attacks. However, it does not detect the presence of the last attack. For this last attack, we can observe that the control input $u(t)$ and the plant state $x_p(t)$ are very slightly affected by the attack signals which allow the attack to escape the detection. Second, the time before detection that is the delay of detection after the occurrence of the safety violation can be studied. Here, the detection occurs after 16.5 s for the first attack and 30.3 s for the second attack. Both statements are likely because the proposed approach is based on ellipsoidal bounds of the reachable sets, meaning that there exists an approximation error. So, at the time before the safety violation, the control input affected by the attacks are most likely inside $\mathcal{R}_u(\infty)$, and thus inside the input ellipsoidal bound $\mathcal{E}_u(R)$. Before the safety violation, the control input is outside $\mathcal{R}_u(\infty)$, but inside $\mathcal{E}_u(R)$. This means that the control input will drive the plant outside the safe set but it cannot be detected by the proposed approach. For the last attack, the control input is most likely operating in this zone, i.e. $u(t) \notin \mathcal{R}_u(\infty)$ and $u(t) \in \mathcal{E}_u(R)$. However, for the first two attacks, the control input at the time detection crosses the boundary of the input ellipsoidal bound leading to the detection. This limitation could be overcome first by using other sets than ellipsoidal sets that often lead to a large approximation error, and second by considering the time evolution of the reachable set instead of using the asymptotic reachable set.

## 6. CONCLUSION

We have proposed a set-theoretic method to design a detection scheme to reveal the presence of stealthy sensor injection attacks by only monitoring the control input to the plant. The novelty of this work lies in monitoring the control input to reveal attacks evading traditional fault-detectors while pushing the plant into critical states. Some limitations have been highlighted that will be the subject of future research works. First, we will explore the possibility of computing an input ellipsoidal bound that will evolve with time to cope with the main limitation of the approach. Second, sets other than ellipsoidal sets will be considered to reduce the error of the outer-approximation of the reachable sets.

## REFERENCES


Azzam, M., Pasquale, L., Provan, G., and Nuseibeh, B. (2022). Efficient predictive monitoring of linear time-invariant systems under stealthy attacks. *IEEE Transactions on Control Systems Technology*, 1–13.

Boyd, S., El Ghaoui, L., Feron, E., and Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory*, volume 15. SIAM.

Cardenas, A.A., Amin, S., Sinopoli, B., Perrig, A., and Sastry, S. (2009). Challenges for securing cyber physical systems. *Proc. 1St Workshop Cyber-Phys. Syst. Security DHS*.

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

Escudero, C., Massioni, P., Zamaï, E., and Raison, B. (2022a). Analysis, prevention, and feasibility assessment of stealthy ageing attacks on dynamical systems. *IET Control Theory & Applications*, 16(4), 381–397.

Escudero, C., Murguia, C., Massioni, P., and Zamaï, E. (2022b). Enforcing safety under actuator injection attacks through input filtering. In *2022 European Control Conference (ECC)*, 1521–1528.

Firoozjaei, M.D., Mahmoudyar, N., Baseri, Y., and Ghorbani, A.A. (2022). An evaluation framework for industrial control system cyber incidents. *International Journal of Critical Infrastructure Protection*, 36, 100487.

Hu, Y., Li, H., Yang, H., Sun, Y., Sun, L., and Wang, Z. (2019). Detecting stealthy attacks against industrial control systems based on residual skewness analysis. *Journal on Wireless Communications and Networking*.

Iqbal, M., Butt, Q.R., and Bhatti, A.I. (2007). Linear model based diagnostic framework of three tank system. In *11th WSEAS International Conference on Systems*.

Kurzhanski, A.B. and Varaiya, P. (2000). Ellipsoidal techniques for reachability analysis: internal approximation. *Systems & control letters*, 41(3), 201–211.

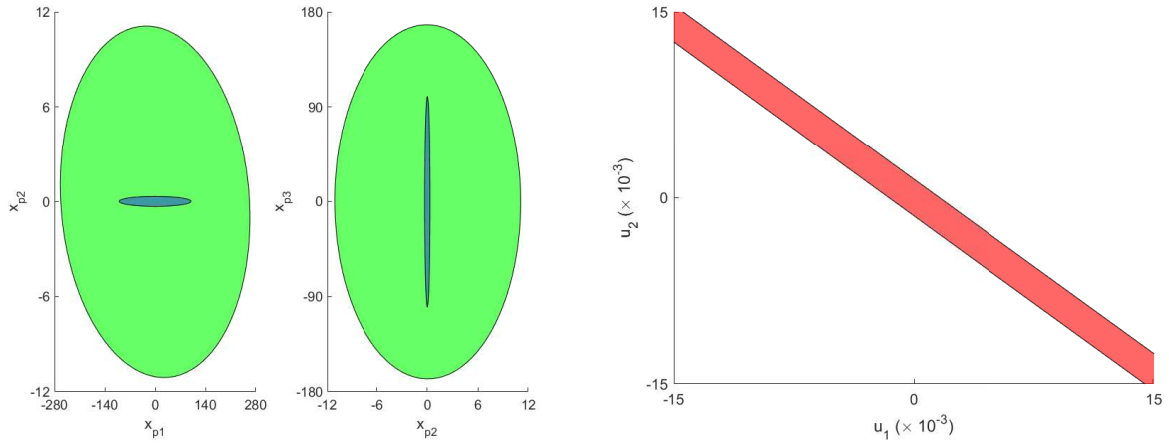


Fig. 2. (Left-hand side) Projection of the state ellipsoidal bound $\mathcal{E}_\zeta(Q)$ onto the $x_p$-hyperplane (green fill), safe set $\mathcal{E}_s(\Psi, \bar{\psi})$ (blue fill) - (Right-hand side) Input ellipsoidal bound $\mathcal{E}_u(R)$ zoomed in (red fill)

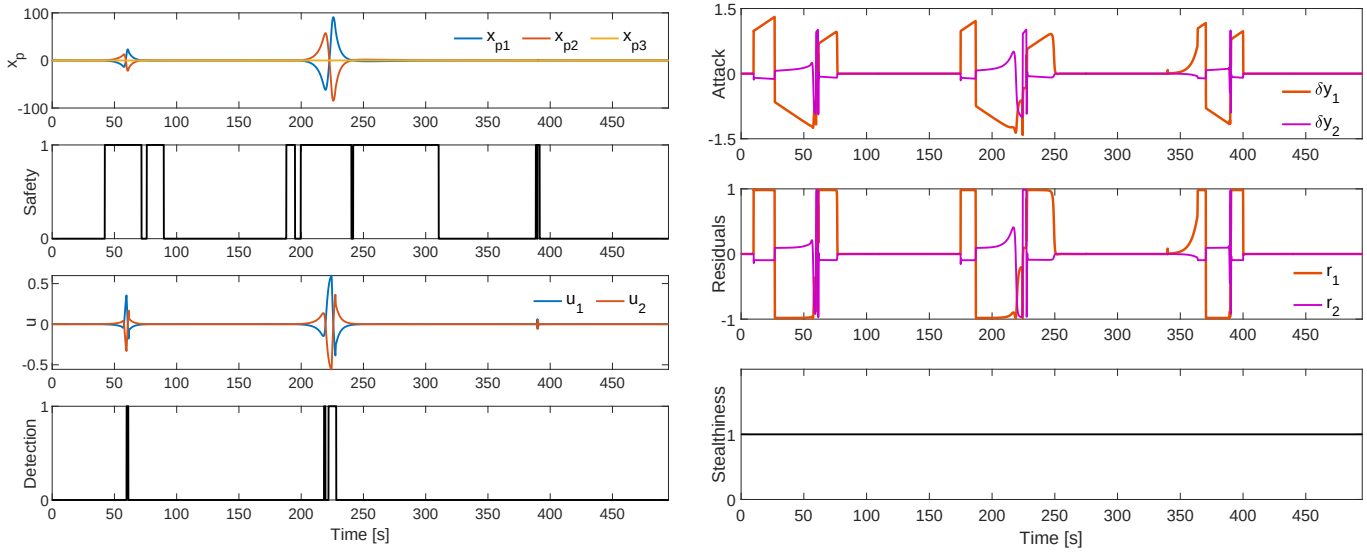


Fig. 3. Stealthy sensor injection attacks $\delta y$ on the closed-loop system: plant states $x_p(t)$ and safety check (top left corner), control input $u(t)$ and detection result (bottom left corner), attack signals $\delta y(t)$ (top right corner), residuals $r(t)$ and stealthiness check (bottom right corner)


Lee, E.A. (2008). Cyber physical systems: Design challenges. In *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, 363–369.

Li, J., Wang, Z., Shen, Y., and Xie, L. (2022). Security synthesis for cyber-physical systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 1–11.

Liu, H., Niu, B., and Qin, J. (2021). Reachability analysis for linear discrete-time systems under stealthy cyber attacks. *IEEE Transactions on Automatic Control*, 66(9), 4444–4451.

Milošević, J., Sandberg, H., and Johansson, K.H. (2020). Estimating the impact of cyber-attack strategies for stochastic networked control systems. *IEEE Transactions on Control of Network Systems*, 7(2), 747–757.

Murguia, C., Shames, I., Ruths, J., and Nešić, D. (2020). Security metrics and synthesis of secure control systems. *Automatica*, 115, 108757.

Sandberg, H., Gupta, V., and Johansson, K.H. (2022). Secure networked control systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1), null.

Teixeira, A., Pérez, D., Sandberg, H., and Johansson, K.H. (2012). Attack models and scenarios for networked control systems. In *Proceedings of the 1st International Conference on High Confidence Networked Systems*, HiCoNS '12, 55–64. ACM, New York, NY, USA.

Yang, T., Murguia, C., Kuijper, M., and Nešić, D. (2021). An unknown input multiobserver approach for estimation and control under adversarial attacks. *IEEE Transactions on Control of Network Systems*, 8(1), 475–486.

Zhang, Q., Liu, K., Pang, Z., Xia, Y., and Liu, T. (2022). Reachability analysis of cyber-physical systems under stealthy attacks. *IEEE Transactions on Cybernetics*, 52(6), 4926–4934.