# A supervisor-based control architecture for constrained cyber-physical systems subject to network attacks

Walter Lucia* Giuseppe Franzè** Bruno Sinopoli***

***Abstract*—In this paper, a novel control architecture for constrained networked control systems is developed with the aim to mitigate the effects of cyber-attacks occurring on the communication channels. By resorting to set-theoretic concepts, physical watermarking ideas and basic cyber-security tools, the proposed supervisor-based control scheme is capable of detecting and mitigating False Data Injection and Denial of Service attacks affecting the normal dynamical evolution of the regulated system. As one of its main merits, the proposed solution guarantees constraint fulfillment, and uniform ultimate boundedness of the regulated system despite any admissible attack realization. Simulation studies are presented to show the capability of the proposed framework while facing different attacks.**

## I. Introduction

Recent advances in sensing, communication and computing have open the door to the deployment of large-scale networks of sensors and actuators that allow fine-grain monitoring and control of a multitude of physical processes and infrastructures. The appellation used by field experts for these paradigms is "Cyber-Physical Systems (CPS) because the dynamics among computers, networking media/resources and physical systems interact in a way that multi-disciplinary technologies (embedded systems, computers, communications and controls) are required to accomplish prescribed missions. Moreover, they are expected to play a significant role in the design and development of future engineering applications such as smart grids, transportation systems, nuclear plants and smart factories [1]. As a consequence, the analysis of security issues has gained an increasing attention from a control perspective, see [2]–[5], and references therein. In this context, a comprehensive classification of the most relevant cyber-attacks (Denial of Service (DoS), False Data Injection (FDI), replay and zero-dynamics), as well as their impact on the CPS security can be found in [6], while a comprehensive survey on the current state-of-the-art can be found in [7].

In the literature, several anomaly/attack detectors have been proposed to detect the presence of cyber-attacks affecting the communication channels in Networked Control Systems (NCSs) (see Fig. 1). The solutions can be classified into two main classes, namely passive and active. Passive schemes [8], [9] attempt to detect malicious activities by exploiting the information extracted from a set of physical observations and without modifying the structure of the underlying control systems. Although appealing, these methods become ineffective under advanced malicious attacks [10], see, e.g. the optimal stealthy attack detailed in [3]. Conversely, active intrusion techniques aim at detecting stealthy attacks by actively manipulating the control system components and/or transmitted actuation and sensor data. Watermarked command signals have been proposed in [11] to detect stealthy steady-state replay attacks; in [2], [5] watermarked sensor measurements, and sensor coding schemes have been used to prevent stealthy measurement attacks; in [12], [13] auxiliary systems, known as moving-targets, are added in the plant-side of the networked control system to avoid coordinated stealthy covert attacks otherwise undetectable by any detector located on the controller side [14].

Cyber-attack countermeasures and resilient control strategies have received increasing attention in the last decade. In [15], the authors have highlighted the limitations of existing fault-tolerant control schemes to deal with cyber-attacks. In [16], [17], reconfiguration algorithms have been designed with the aim to mitigate the undesired effects of cyber-attacks affecting the class of NCSs. Control solutions against DoS and resource-constrained attackers have been analyzed in [18]–[21], while replay and packet scheduling attacks have been discussed in [22], [23]. In [24], an unconstrained model predictive control (MPC) algorithm has been developed to compensate deception attack occurrences, which give rise to time-varying network delays, packet disorders and data losses. In [25], the secure control problem of cyber-physical systems is recast in terms of a zero-sum stochastic game. Such an approach is used to design a switching policy for unconstrained linear systems under different cyber-attacks. Finally, in [26], [27], adaptive controllers for unconstrained and disturbance-free linear systems are proposed to guarantee uniform ultimate boundedness of the regulated system under sensor and actuator attacks. All these contributions share as a common denominator, the lack of a robust and resilient control framework for networked system configurations capable to jointly take care of state/input constraints, disturbances and cyber-attack occurrences on the communication medium.

*Walter Lucia is with the Concordia Institute for Information Systems Engineering (CIISE), Concordia University, Montreal, QC, H3G 1M8, CANADA walter.lucia@concordia.ca

**Giuseppe Franzè is with DIMES department, University of Calabria, Via Pietro Bucci, Cubo 42-C, Rende (CS), 87036, ITALY giuseppe.franze@unical.it

***Bruno Sinopoli is with the Electrical and Systems Engineering department, Washington University in St Louis, St. Louis, MO 63130, USA, bsinopoli@wustl.edu

### A. Paper Contribution

In this paper, a supervisor-based architecture for the resilient control of constrained systems under DoS and FDI attacks is developed by taking advantage of set-theoretic receding horizon control ideas [28] and watermarking arguments [11].

To the best of the authors' knowledge, the proposed control architecture addresses two important issues that the existing state-of-the-art has not yet addressed simultaneously: state and input constraints fulfillment, and resiliency under arbitrary DoS and FDI attack occurrences. In particular, in [19]–[21], FDI attacks and constraints are not considered; in [22], state and input constraints are handled but the proposed solution is effective only against a very specific FDI attack (i.e., replay attacks). Finally, in [26], an unconstrained setup is assumed and the FDI attack occurrences are limited to be a function of the state of the system. The main novelties of the proposed approach can be summarized as follows:

- The proposed control architecture represents one of the first attempts to deal with CPS subject to: *(i)* state and input constraints, *(ii)* bounded disturbances and *(ii)* FDI and DoS attacks on both the actuation and measurement channels.
- A novel active detection mechanism acting on both plant and controller sides has been derived to detect intelligent FDI attacks by jointly resorting to set-theoretic model predictive control and watermarking ideas. One of its novel features is that the watermarking signal is not superimposed on the control signal [11], but it is instead embedded into the controller logic design.
- The concept of robust one-step controllable set [28] has been extended to deal with DoS/FDI attack occurrences making the communication channel unreliable, and as a consequence, forcing the system to operate in an open-loop fashion. The latter has been addressed by introducing the concept of robust $\tau-$step controllable sets that allows to compute control inputs that, whenever necessary, can be safely and constantly applied to the plant for $\tau-$ step, so avoiding constraints violations.

The paper is organized as follows. In section II, some basic definitions used along the manuscript are introduced, and the set-theoretic receding horizon control paradigm is revised; in section III, the proposed supervised control architecture is introduced, and the problem formulation formally stated; in section IV, the proposed set-based detector is developed, and its properties proved; in section V, first the set-theoretic resilient controller and the supervisor module are designed, then a computational algorithm is detailed and uniform ultimately boundedness of the closed-loop system proved; finally, section VI shows a numerical simulation example highlighting the capabilities of the proposed solution.

## II. Preliminaries and Notations

We shall consider the following class of discrete-time linear time-invariant (LTI) systems

$$x(t+1) \quad = \quad Ax(t) + Bu(t) + B_d d(t) \tag{1}$$

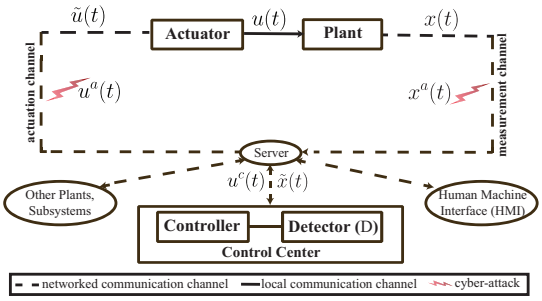


Fig. 1. Networked Control System (NCS) subject to cyber-attacks on the networked communication channels.

where $x(t) \in \mathbb{R}^n$ is the plant state space vector, $t \in \mathbb{Z}_+ := \{0, 1, ...\}$ the sampling time instant, $u(t) \in \mathbb{R}^m$ the control input vector, and $d(t) \in \mathcal{D} \subset \mathbb{R}^d$ an exogenous plant disturbance bounded by the compact set $\mathcal{D}$. Moreover, (1) is subject to state and input set-membership constraints:

$$u(t) \in \mathcal{U}, \quad x(t) \in \mathcal{X}, \ \forall t \geq 0, \tag{2}$$

where $\mathcal{U}$ and $\mathcal{X}$ are compact subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, with the origins as interior points.

*Definition 1:* Let $S \subset \mathbb{R}^n$ be a neighborhood region of the origin. The autonomous system $x(t+1) = Ax(t) + B_d d(t)$ is said to be Uniformly Ultimately Bounded (UUB) in $S$ if for all $\mu > 0$ there exists $T(\mu) > 0$ such that $\forall \|x(0)\| \leq \mu \rightarrow x(t) \in S \ \forall d(t) \in \mathcal{D}$ and $\forall\, t \geq T(\mu)$ [28].

*Definition 2:* A set $\mathcal{T} \subseteq \mathcal{X}$ is said to be a Robust Control Invariant (RCI) set for (1) under the disturbance $d(t) \in \mathcal{D}$ and constraints (2) if there exists a control law $u := f(x(t)) \in \mathcal{U}$ such that $\forall\, x(0) \in \mathcal{T} \rightarrow Ax(t) + Bf(x(t)) + B_d d(t) \in \mathcal{T}, \ \forall d(t) \in \mathcal{D}, \forall t \in \mathbb{Z}_+$ [29]. □

*Definition 3:* Given the sets $\mathcal{A}$, $\mathcal{E} \subset \mathbb{R}^n$, $\mathcal{A} \oplus \mathcal{E} := \{a + e : a \in \mathcal{A}, e \in \mathcal{E}\}$ is the *Minkowski Set Sum* and $\mathcal{A} \sim \mathcal{E} := \{a \in \mathcal{A} : a + e \in \mathcal{A}, \forall e \in \mathcal{E}\}$ the *Pontryagin-Minkowski Set Difference*. □

*Definition 4:* Given a polyhedron $\mathcal{P} \subset \mathbb{R}^p$ and a matrix $T \in \mathbb{R}^{q \times p}$, the affine map (polyhedron) of $\mathcal{P}$ along $T$ is:

$$\mathcal{Q} := \{y \in \mathbb{R}^q \ : y = Tx, \ x \in \mathcal{P}\}$$

*Definition 5:* Given the constrained system (1)-(2) and a positive scalar $\tau$, a state-feedback control $u(t) = K_{IOD} x(t - \tau(t))$ is defined Independent-of-Delay (IOD) if the closed-loop system is robustly stable and satisfies (2) $\forall d(t) \in \mathcal{D}$, $\forall t \geq 0$ and $\forall \tau(t) \leq \tau$ [30]. □

*Definition 6:* Consider the NCS in Fig. 1 and a networked communication channel, namely $ch$. An attacker has "*disclosure*" resources on $ch$ if it can read the transmitted data. An attacker has "*disruptive*" resources on $ch$ if it can change the transmitted data [6].

*Definition 7:* Consider the NCS in Fig. 1 and the anomaly/attack detector module $(D)$. In the sequel, a cyber-attack against the networked control system operations is said stealthy or undetectable if it will never trigger an anomaly on $D$.

### A. Set-theoretic receding horizon control scheme (ST-RHC)

The receding horizon control scheme developed in [31] and based on the philosophy proposed in [32] is in this subsection summarized.

The regulation problem for constrained LTI systems (1)-(2) is addressed by resorting to a dual-mode receding-horizon control strategy based on a family of robust one-step controllable regions. Such sets are off-line computed and on-line exploited as target sets for the one-step state predictions. This translates into the following algorithm:

——————— *(ST-RHC)* algorithm ———————

*Off-line -*

1) Compute a stabilizing state-feedback control law $u^0(\cdot) = f^0(x(\cdot))$ complying with (2) and the associated RCI region $\mathcal{T}^0$;
2) Starting from $\mathcal{T}^0$, recursively compute a sequence of $N$ robust one-step controllable sets $\{\mathcal{T}^i\}_{i=1}^N$ according to the following definition [28]:

$$\begin{aligned} \mathcal{T}^0 &:= \mathcal{T} \\ \mathcal{T}^i &:= \{x \in \mathcal{X} : \exists\, u \in \mathcal{U}\ s.t. \\ &\qquad Ax + Bu + B_d d \in \mathcal{T}^{i-1}, \forall d \in \mathcal{D}\}, \\ &\qquad i = 1, \dots, N. \\ &= \{x \in \mathcal{X} : \exists\, u \in \mathcal{U}\ s.t.\ Ax + Bu \in \tilde{\mathcal{T}}^{i-1}\}, \\ &\qquad i = 1, \dots, N. \end{aligned} \tag{3}$$

where $\tilde{\mathcal{T}}^{i-1} := \mathcal{T}^{i-1} \sim B_d\mathcal{D}$.

*On-line -*

1) Find $i(t) := \min\{i : x(t) \in \mathcal{T}^i\}$
2) If $i(t) = 0$ then

$$u(t) = f^0(x(t)) \tag{4}$$

3) Else solve the following quadratic programming (QP) problem:

$$\begin{aligned} u(t) = \arg\min_u J(t, x(t), u) \quad s.t. \\ Ax(t) + Bu \in \tilde{\mathcal{T}}^{i(t)-1},\ u \in \mathcal{U} \end{aligned} \tag{5}$$

with $J(t, x(t), u)$ a convex cost function

———————————————————————————

***Remark** 1:* The objective of the *ST-RHC* controller is to drive the regulated state trajectory within the terminal region $\mathcal{T}^0$ in a-priori defined number $N > 0$ of steps (guaranteed by construction) and irrespective of the used cost function $J(t, x(t), u)$, see [31] for a detailed discussion. The latter makes $J(t, x(t), u)$ a degree of freedom of the strategy that, therefore, can be arbitrarily changed at each time instant without affecting the feasibility of (5). □

***Remark** 2:* It is worth to remark that within the same control framework, bounded state measurement noises (e.g. $y(t) = x(t) + \eta(t)$, with $\eta(t)$ a bounded random variable) can be straightforwardly considered by properly customizing the definition of robust one-step controllable sets (3) as done e.g. in [33]. Since the occurrence of measurement noises does not modify the *modus operandi* of the proposed scheme, here it has omitted for improving the clarity of next developments.

## III. Problem formuation

We consider an operating scenario where malicious agents can alter the networked communications in a NCS (Fig. 1) causing Denials of Service and False Data Injections. FDI and DoS attacks on the communication channels are modeled as follows:

- *controller-to-actuator link:*

$$\tilde{u}(t) := u^c(t) + u^a(t)\ \ (\textit{FDI}),\ \ \tilde{u}(t) := \emptyset\ \ (\textit{DoS}) \tag{6}$$

- *sensor-to-controller link:*

$$\tilde{x}(t) := x(t) + x^a(t)\ \ (\textit{FDI}),\ \ \tilde{x}(t) := \emptyset\ \ (\textit{DoS}) \tag{7}$$

where $u^a(t) \in \mathbb{R}^m$ and $x^a(t) \in \mathbb{R}^n$ are unknown and unbounded malicious signals, while $\tilde{u}(t) \in \mathbb{R}^m$ and $\tilde{x}(t) \in \mathbb{R}^n$ account for the resulting corrupted control signals and state measurements, respectively.

The considered control problem can be stated as follows:

**Resilient Control of NC-CPSs subject to cyber-attacks (RC-NC-CPS)** - *Consider the control architecture of Fig. 1. Given the NC-CPS plant model (1)-(2) subject to FDI attacks (6)-(7), design*

- **(P1)** *An active robust anomaly detector $D$ capable to discover cyber-attack occurrences;*
- **(P2)** *A control strategy $u(\cdot) = f(\tilde{x}(\cdot))$ such that the closed-loop trajectory is Uniformly Ultimately Bounded and the prescribed constraints are fulfilled regardless of any admissible DoS or FDI attack scenario. Moreover, if $u^a(t) \equiv 0$, $x^a(t) \equiv 0$ (attack free scenario) and $d(t) \equiv 0$ (disturbance free scenario) $\forall t \geq \bar{t}$, then the regulated system is asymptotically stable.*

For solvability reasons, which will be clarified in Section V, the following assumptions are made:

*Assumption 1:* A guaranteed attack-free communication between the controller and the plant can be reestablished in at most $T_{new}$ time steps.

*Assumption 2:* The time interval $T_{viol}$ required to breach the communication protocol is non-vanishing, i.e. $T_{viol} \geq T_{new}$.

*Assumption 3: Communication latency is negligible.*

***Remark** 3:* First, notice that *Assumptions* 2-3 are standard when SCADA infrastructures are of interest. First, it is important to notice that SCADA systems make use of real-time network (fieldbus) protocols, e.g. DNP3, Modbus, Profibus, Profinet, Cip, over Serial or Ethernet communication channels, that are reliable in the absence of attacks (*Assumption* 3), although not equipped with authentication and encryption mechanisms. Moreover, basic security countermeasures (ethernet switches or routers protected by a firewall) are always in place to avoid commonplace intrusions. As a consequence, a non-vanishing time interval $T_{viol}$ is required to breach any communication protocol successfully (*Assumption* 2).

As the *Assumption* 1 is concerned, it appears also reasonable in virtue of the following arguments. In the NCSs field, a single communication channel is traditionally adopted between the plant side and the control center. Such a simplified framework has to be updated when the CPSs structure becomes so complex to provide reliable and timely communications

This article has been accepted for publication in IEEE Transactions on Control of Network Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3182098



[34]. Then, multi-channels technologies are exploited in order to increase communication reliability and resiliency against severe circumstances (disasters or cyber-attacks), see e.g. [35]. In this case, it is quite realistic to assume that a resource-limited adversary is not capable of compromising all the communication links by means of a single attack. Along these lines, notice that there exist several approaches to efficiently re-route the communication between two entities in a multi-path channel, see [36] and references therein for details. □

In what follows, the **RC-NC-CPS** problem will be addressed by properly customizing the set-theoretic ideas of the RHC scheme presented in Section II-A. The proposed Networked Constrained Cyber-Physical Systems (NC-CPS) is depicted in Fig. 2 and it includes the following key modules:

- a **detector** module $D$ (designed in Section IV) that is in charge of robustly detecting cyber-attack occurrences without generating false alarms;
- a **supervisor** module (designed in Section V-B) that checks the admissibility and correctness of the received commands $\tilde{u}(t)$ (Pre-Check unit task) and plant states $x(t)$ (Post-Check unit task). The output of the supervisor, namely $w(t)$, is sent to the smart actuator;
- a **smart actuator** module (designed in Section V-C) that, given $w(t)$, applies either the received command $\tilde{u}(t)$ or a previously stored admissible input (Section V-A).

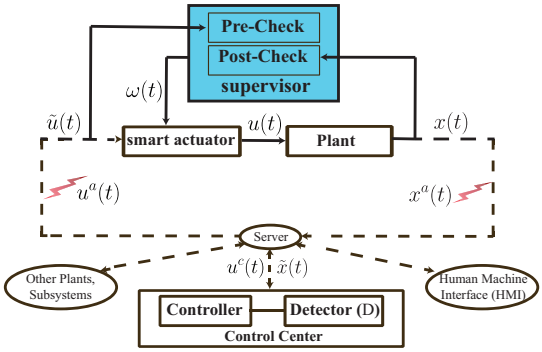


Fig. 2. Reliable Control Architecture Against Network Attacks

## IV. Set-theoretic characterization and detection of attacks

In this section, a set-based robust detector for FDI attacks is first introduced and then customized to obtain an active watermarked anomaly detector. Notice that the detection of DoS attacks is not considered in the sequel because if an empty packet ($\tilde{u}(t) = \emptyset$ or $\tilde{x}(t) = \emptyset$) is received then it is trivial that a DoS has occurred.

### A. Basic Set-Based Passive Anomaly Detector

Let $x$ and $u^c$ be the current state and command input, respectively. Then, the expected robust one-step prediction set $\mathcal{X}^+(x, u^c) \neq \emptyset$ (see Fig. 3) can be defined as follows:

$$\mathcal{X}^+(x,u^c) := \{x^+ \in \mathbb{R}^n : x^+ = Ax + Bu^c + B_d d, \forall\, d \in \mathcal{D}\} \quad (8)$$

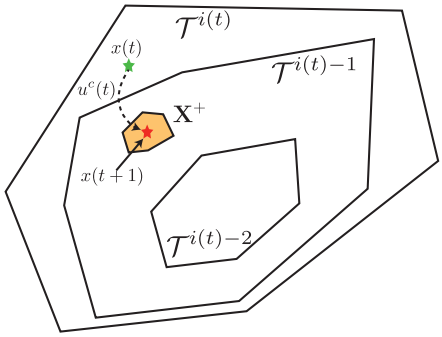


Fig. 3. State evolutions under the action of $u^c(t)$ computed via (5). The orange region defines the one-step prediction set $\mathcal{X}^+$

By leveraging the robust one-step prediction set $\mathcal{X}^+$ (8), the following passive and robust binary attack detection rule can be defined:

$$D(\tilde{x}(t)) := \begin{cases} \textit{anomaly} & \text{if } \tilde{x}(t) \notin \mathcal{X}^+(\tilde{x}(t-1), u^c(t-1)) \\ \textit{normal} & \text{otherwise} \end{cases} \quad (9)$$

### B. Enhanced Active Anomaly Detector

Since the proposed detector (9) is passive, it is not possible to detect advanced coordinated FDI attacks, see [8], [10].

To mitigate such a drawback, we enhance the basic detector $D$ with an active watermarking-based effect [11]. Differently from existing implementations where a noisy signal is superimposed on the optimal action, here we take advantage of the properties of the set-theoretic controller (5) where the cost function $J(t, x(t), u)$ can be arbitrarily selected during the online operations, see *Remark* 1.

In particular, such a degree-of-freedom can be exploited to prevent the attacker's chance of emulating the ST-RHC control actions. This translates into the following steps:

- offline - define a set of $1 < L < \infty$ cost functions and state-feedback control laws compatible with $\mathcal{T}^0$,

$$\mathcal{J} = \{J_k(\tilde{x}(t), u)\}_{k=1}^{L}, \ \mathcal{F}^0 = \{f_k^0(\tilde{x}(t))\}_{k=1}^{L} \quad (10)$$

- online - $(\forall\, t)$, use an uniformly distributed random function

$$j(t) : \mathbb{Z}_+ \to [1, \ldots, L] \quad (11)$$

to select the cost function in (5) or the terminal state-feedback control law in (4).

As a consequence, we have that: *(i)* the detection scheme becomes active, (ii) steady-state replay-attacks are no longer stealthy (see [11]), (iii) the control logic is randomized and unknown to the attacker.

The following proposition characterizes the class of FDI attacks undetectable by the watermarked detector $D$.

*Proposition 1: Consider the NC-CPS model (1)-(2), the anomaly detector (9), a set of $L > 1$ penalizing functions $\mathcal{J}$, $L$ state-feedback control law $\mathcal{F}^0$, and an uniformly random function $j(t)$, defined as in (11), and unknown to the attacker. If $\forall\, t$ the command $u^c(t)$ is obtained from (4) or (5) with $f_{j(t)}^0(\tilde{x}(t)) \in \mathcal{F}^0$, $J_{j(t)}(\tilde{x}(t), u) \in \mathcal{J}$, then FDI attacks (6)-(7)*

This article has been accepted for publication in IEEE Transactions on Control of Network Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3182098



*can be stealthy if and only if the attacker has the following assets:*

- *knowledge of the plant model (1)-(2);*
- *disruptive resources on the data transmitted on the actuation ($u^c(t)$) and measurement channels ($x(t)$);*
- *disclosure resources on the data transmitted on the actuation channel ($u^c(t)$).*

*Proof -* (*sufficient condition*): If the attacker has plant model knowledge (1)-(2), disruptive resources on the actuation and measurement channel and disclosure resources on the actuation channel, then the following coordinated constrained attack can be performed [14]: first, the adversary injects on the controller-to-actuator channel an additive admissible perturbation $u^a(t-1)$ such that

$$\tilde{u}(t-1) = \hat{u}^c(t-1) + u^a(t-1) \in \mathcal{U}, \tag{12}$$

then it removes the effect of such an attack from the measurement channel by injecting

$$x^a(t) = -\sum_{j=0}^{t-1}(A^j B u^a(t-1-j)) \tag{13}$$

According to the detection rule (9), such attack is, by construction undetectable, i.e. $\tilde{x}(t) \in D(\tilde{x}(t)), \forall t$.

(*necessary condition*): If the attacker is not aware of the plant matrices $A, B$ and disturbances set $\mathcal{D}$, the attack cannot determine $\mathcal{X}^+(x(t-1), u^c(t-1))$ and it cannot generate an attack capable of satisfying the detection rules defined in $\mathcal{D}$, i.e. $\tilde{x}(t) \in \mathcal{X}^+(x(t-1), u^c(t-1)), \forall t$. On the other hand, the constrained attack (12)-(13) trivially cannot be performed. Moreover, if the attacker is aware of (1)-(2), but it hasn't disruptive resources on one of the communication links, then a coordinated attack cannot be performed. In this case, the only possibility for the attacker to be stealthy is to inject a small perturbation in either the actuation or measurement channels. However, from one hand, such an injection is not guaranteed to be stealthy, while, on the other hand, the effect of such an attack is nullified by the robust nature of the ST-RHC controller. Finally, if the attacker has all the required resources except disclosure resources on the actuation channel, then the attacker is not aware of the randomized controller action $u^c(j(t))$ and it cannot determine the expected one-step evolution $\mathcal{X}^+(\tilde{x}(t), u^c(j(t))), \forall t$ (see Fig.4). Moreover, the constrained attack (12)-(13) is not guaranteed to be stealthy because linear arguments cannot be exploited as in (13). □

Note that the anomaly detection rule (9) takes into account, by construction, the worst-case realization of the disturbance $d \in \mathcal{D}$. As a consequence, the absence of false positives is ensured. On the other hand, as for any other anomaly detector for CPSs [37], the absence of false negatives cannot be generally guaranteed. Therefore, if an attack is not instantaneously detected, then the safety of the system might be potentially at risk. To address such a drawback, the next section will design a resilient and secure control strategy that ensures constraint satisfaction and UUB of the closed-loop state trajectory regardless of any admissible attack scenario.

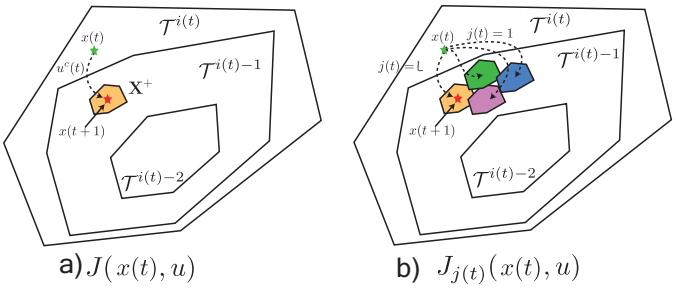


Fig. 4. One-step ahead state evolutions: (a) fixed cost function $J(x(t), u)$ vs (b) randomly chosen cost function $J_{j(t)(x(t),u)}$

## V. Cyber-Physical countermeasures for resilient and secure control

If an attack has been detected, the following countermeasures should be adopted to recover an attack-free operating scenario: 1) disconnect *sensor-to-controller* and *controller-to-actuator* communication links; 2) reestablish secure channels. The main consequence of such actions is that for a given time interval, namely $T_{new}$, updated state measurements and command inputs are not available at the controller and actuator sides, respectively. Therefore, the following question assumes a key relevance:

*How can safety requirements $x(t) \in \mathcal{X}$, $u(t) \in \mathcal{U}$ be satisfied while communication channels are out of order for $T_{new}$ time instants?*

In the resilient architecture of Fig. 2, the supervisor is in charge of dealing with this critical scenario. Essentially, the idea is to take advantage of the features of the *ST-RHC* strategy to formally define an admissible, though not optimal, command input that can be consecutively applied for a given time interval. To such hand, it will be shown that two units, namely Pre-Check and Post-Check, are needed to detect anomalies on the received commands $\tilde{u}(t)$ and state measurements $x(t)$.

### A. Resilient command input

Here, we provide the conditions under which a constant input can be repeatedly applied without compromising constraints satisfaction and feasibility retention.

*Lemma 1: Given the constrained system (1)-(2) and a positive integer $\tau$, a sequence of $\tau$-steps controllable regions computed according to the following set recursions*

$$\begin{aligned}\mathcal{T}_\tau^0 &:= \mathcal{T} \\ \mathcal{T}_\tau^i &:= \bigcap_{k=1}^{\tau} \left\{ x \in \mathcal{X} : \exists\, u \in \mathcal{U} : A^k x + \left( \sum_{j=0}^{k-1} A^j B \right) u \in \tilde{\mathcal{T}}_\tau^{i-1} \right\}\end{aligned} \tag{14}$$

*where $\mathcal{T}$ is a RCI region for (1)-(2) equipped with the (IOD) state-feedback control law $f^0(x(\cdot)) := K_{IOD} x(t-\tau)$, $K_{IOD} \in \mathbb{R}^{m \times n}$, and*

$$\tilde{\mathcal{T}}_\tau^i = \mathcal{T}_\tau^i \sim \bigcup_{k=1}^{\tau} A^{k-1} B_d \mathcal{D}_x, \;\; i \geq 1 \tag{15}$$

*guarantees that for any $x(t) \in \mathcal{T}_\tau^i$ there exists an admissible command $u^c \in \mathcal{U}$ that can be constantly applied for $\tau$ time instants to (1) without constraint violations.*

*Proof -* If the current state $x(t)$ belongs to $\mathcal{T}_\tau^i$, $i \geq 1$, then there exists a constant command $u^c \in \mathcal{U}$ that produces a $\tau-$step ahead state evolution robustly confined into the region $\mathcal{T}_\tau^i \subseteq \mathcal{X}$. Indeed, the sets $\tilde{\mathcal{T}}_\tau^i$ are built-up to take care of any disturbance occurrence along the state predictions, see (15). Conversely, when $x(t) \in \mathcal{T}_\tau^0$, by resorting to the ideas argued in [38], the state-feedback law $f^0(x(t))$ is feasible and applicable for $\tau$ consecutive steps from the last available state measurement. □

Then, by describing the $\tau-$ steps controllable regions in the extended space $(x,\, u)$ :

$$\Xi_\tau^i := \bigcap_{k=1}^{\tau} \left\{ (x,\, u) \in \mathcal{X} \times \mathcal{U} : A^k x + \left( \sum_{j=0}^{k-1} A^j B \right) u \in \tilde{\mathcal{T}}_\tau^{i-1} \right\} \tag{16}$$

and by exploiting the projection operator:

$$\mathcal{T}_\tau^i = Proj_x\{\Xi_\tau^i\}, \quad \mathcal{U}_\tau^i = Proj_u\{\Xi_\tau^i\}, \tag{17}$$

the following result comes out.

*Proposition 2: Let $T_{new} > 0$, $N > 0$, $\{\Xi_{T_{new}}^i\}_{i=1}^N$, $\{\mathcal{U}_{T_{new}}^i\}_{i=1}^N$, $\{\mathcal{T}_{T_{new}}^i\}_{i=1}^N$ be given. If $x(t) \in \mathcal{T}_{T_{new}}^{i(t)}$, $u^a(t) \equiv 0$, $y^a(t) \equiv 0$ (attack free scenario), then the control input $u^c(t)$ computed by means of the following convex optimization problem*

$$\begin{aligned} u^c(t) &= \arg\min_u J_{j(t)}(x(t), u) \quad s.t. \\ [x(t),\, u] &\in \Xi_{T_{new}}^i, \quad u \in \mathcal{U}_{T_{new}}^i \end{aligned} \tag{18}$$

*and consecutively applied to (1) for $T_{new}$ time instants, guarantees: i) constraints fulfillment; ii) state trajectory confinement, i.e. $x(t+k) \in \mathcal{T}_{T_{new}}^{i(t)-1}$, $\forall k = 1, \ldots, T_{new}$, despite any realization of $d(\cdot) \in \mathcal{D}$ and any choice of $J_{j(t)}(x(t), u) \in \mathcal{J}$.*

*Proof -* If $x(t) \in \mathcal{T}_{T_{new}}^{i(t)}$ then there always exists an admissible command input $u^c(t)$ that is solution of the optimization (18). Such a command, if consecutively applied to (1) for $T_{new}$ steps, leads to the following state evolution for $1 \leq t \leq T_{new}$

$$x(t+k) = x(t+k)^{u^c} + x^d(t+k)$$

where

$$\begin{aligned} x^{u^c}(t+k) &:= A^k x(t) + \sum_{j=0}^{k-1} (A^j B) u^c(t) \\ x^d(t+k) &:= \sum_{j=0}^{k-1} (A^j B_d) d(j) \end{aligned}$$

Therefore, in virtue of the Pontryagin-Minkowski set difference recursion (15), if $x^{u^c}(t+k) \in \tilde{\mathcal{T}}_{T_{new}}^{i-1}$ then $x(t+k) \in \mathcal{T}_{T_{new}}^{i-1}$ for any disturbance realization $d(\cdot) \in \mathcal{D}$. □

***Remark 4:*** Notice that the optimization (18) also allows to adequately address DoS occurrences. In fact, if the DoS attack has a time duration less or equal than $T_{new}$ time instants, then the stored command $u_{-1}$ can be repeatedly applied until a new admissible data is received. □

The next two sections will be devoted to deal with operating scenarios where the malicious agent is capable to alter the last command input $\tilde{u}(t)$ received by the Actuator (see Fig. 2). In such a case, the countermeasures of Proposition 2 are no longer valid.

### *B. Pre-Check and Post-Check design*

The aim of these units is to certify that at each time instant $t$ the received command $\tilde{u}(t)$ and state measurement $x(t)$ are "coherent" with the expected set-level $i(t)$ : $\tilde{u}(t) \in \mathcal{U}_{T_{new}}^{i(t)}$ and $x(t) \in \mathcal{T}_{T_{new}}^{i(t)}$. Notice that the sets $\mathcal{U}_{T_{new}}^{i(t)}$ and $\mathcal{T}_{T_{new}}^{i(t)}$ have the same structure of (17) with $\tau \leftarrow T_{new}$ and $i \leftarrow i(t)$. In particular, the following logical rules are exploited:

$$\text{Pre-Check}(i(t)) := \begin{cases} true & \text{if } \tilde{u}(t) \in \mathcal{U}_{T_{new}}^{i(t)} \\ false & \text{otherwise} \end{cases} \tag{19}$$

$$\text{Post-Check}(i(t)) := \begin{cases} true & \text{if } x(t) \in \mathcal{T}_{T_{new}}^{i(t)} \\ false & \text{otherwise} \end{cases} \tag{20}$$

According to the Proposition 2 statement, the NC-CPS state evolution and the applied commands are confined within $\mathcal{T}_{T_{new}}^{i(t)}$ and $\mathcal{U}_{T_{new}}^{i(t)}$, respectively, while the switching index $i(t)$ exhibits a monotonically non-increasing trend. Therefore, if one of the set-membership checks (19)-(20) fails, then the following actions need to be taken to ensure plant safety:

- *(a)* if $\tilde{u}(t) \notin \mathcal{U}_{T_{new}}^{i(t)}$ and $x(t) \in \mathcal{T}_{T_{new}}^{i(t)}$, then $\tilde{u}(t)$ is discarded and a stored input, hereafter named $u_{-1} := \tilde{u}(t-1)$, applied;
- *(b)* if $x(t) \notin \mathcal{T}_{T_{new}}^{i(t)}$, then an harmful command has been applied at the previous time instant $\tilde{u}(t-1)$ and, as a consequence, the stored command $u_{-1}$ is no longer trustworthy. An admissible countermeasure consists in applying at the next time instants the zero input $u(t) \equiv 0_m$ (open-loop mode) and in triggering the detector to reestablish safe communication channels.

Notice that in the scenario *(a)* the prescribed countermeasure can be straightforwardly applied in virtue of Proposition 2, whereas in the scenario *(b)* the prescribed countermeasure can be safely applied only if the zero-input open-loop evolution will remain confined within the controller Domain of Attraction (DoA), i.e.

$$\bigcup_{k=1}^{T_{new}} \underbrace{(A^k \mathcal{T}_{T_{new}}^{i(t)} \oplus \sum_{j=0}^{k-1} A^j B_d \mathcal{D}_x}_{\textit{first term}} \oplus \underbrace{A^{k-1} B \mathcal{U}}_{\textit{second term}}) \subseteq \bigcup_{j=1}^{N} \mathcal{T}_{T_{new}}^j \tag{21}$$

where the *first term* accounts for the autonomous state evolution of (1) and the *second term* for the perturbation arising from the application of a single unknown admissible input. Moreover, we denote with $i_{max} \leq N$ the maximum set level for which the following set-containment holds true

$$\bigcup_{k=1}^{T_{new}} \left( A^k \mathcal{T}_{T_{new}}^{i_{max}} \oplus \sum_{j=0}^{k-1} A^j B_d \mathcal{D}_x \oplus A^{k-1} B \mathcal{U} \right) \subseteq \bigcup_{j=1}^{N_{max}} \mathcal{T}_{T_{new}}^j \tag{22}$$

where the upper bound $N_{max} := \min(N, i_{max} + T_{viol})$ is introduced to deal with the possibility that a new attack

can occur $T_{viol}$ time instants after the recovery phase (see Assumption 2).

The reasoning behind the introduction of (22) relies on the following arguments. When communications are interrupted, the NC-CPS (1) operates in an open-loop fashion under zero-input actions. In such a scenario, the set-containment (22) guarantees that, starting from any initial condition within $\mathcal{T}_{T_{new}}^{i_{max}}$, the resulting $T_{new}-$ steps state evolution is in the worst case confined into $\bigcup_{j=1}^{N_{max}}\{\mathcal{T}_{T_{new}}^{j}\}$. Therefore, feasibility of the underlying RHC strategy is retained and the UUB property of the closed-loop behavior of (1) guaranteed.

***Remark** 5:* Note that the supervisor module cannot be designed to replace the detector $\mathcal{D}(\tilde{x}(t))$. To motivate this statement we need to recall that if $\tilde{u}(t) \in \mathcal{U}_{T_{new}}^{i(t)}$ then the supervisor is not able to understand if $\tilde{u}(t) \neq u^c(t)$. Therefore, it cannot determine the expected robust one-step prediction set $\mathcal{X}^+(x(t), u^c(t))$ as done by $\mathcal{D}(\tilde{x}(t))$. Moreover, $u^c(t)$ cannot be re-computed (on the plant side) from the available state measurement $x(t)$ because the random function $j(t)$, used to compute $u_c(t)$, is unknown. On the other hand, the supervisor module can only check if $[x(t),\, \tilde{u}(t)] \in \Xi_{T_{new}}^{i(t)}$ as prescribed by (16). □

### C. Supervisor and Actuator logics

The supervisor's output map is defined as follows:

$$\begin{gathered}\omega(t) = \eta(x(t), \tilde{u}(t)) \\ \eta(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R}^m \to \{NORMAL,\, SKIP,\, REJECT\}\end{gathered} \tag{23}$$

where $NORMAL$ refers to an attack not detected by neither the Pre-Check nor Post-Check units, $SKIP$ if the last received command input cannot be applied (the case *(a)*) and $REJECT$ if communication channels cannot be reputed "safe" (the case *(b)*). From a computational point of view, the implicit map $\eta(\cdot,\cdot)$ can be described by the following algorithm:

——— **Supervisor Logic** $\eta(x(t), \tilde{u}(t))$ ———

**if Pre-Check**$(i)$==true & **Post-Check**$(i)$==true **then**
  $\omega(t) = NORMAL$,
**else**
  **if Post-Check**$(i)$==true & **Pre-Check**$(i)$==false **then**
    $w(t) = SKIP$; ▷ Attack locally detected
  **else**
    $w(t) = REJECT$ ▷ Attack locally detected
  **end if**
**end if**
Send $\omega(t)$ to the **Actuator**

Conversely, the **Actuator** exploits the following logical rules:

——— **Smart Actuator Logic** ———

**Initialization:** $u_{-1} = u^c(0)$

**if** $\tilde{u}(t) = \emptyset$ **then** ▷ Input data loss (DoS)
  $u(t) = u^{-1}$; ▷ Apply the stored command
**else**
  **if** $\omega(t) == NORMAL$ **then** $u(t) = \tilde{u}(t)$ ▷ Normal
  **else**
    **if** $\omega(t) == SKIP$ **then** $u(t) = u^{-1}$;
    **else** $u(t) = 0$ ▷ REJECT: open-loop mode
    **end if**
  **end if**
**end if**
Apply $u(t)$ to the plant
$u^{-1} \leftarrow u(t)$

### D. The RHC algorithm

In the sequel, a RHC based algorithm, hereafter denoted as $\tau$**-ST-RHC**, is developed.

$\tau$**-ST-RHC** Algorithm

——— **Control Center** ———
(off-line)

**Input:** $\mathcal{T}^0$, $N$, $T_{new}$;
**Output:** $\{\Xi_{T_{new}}^i\}_{i=0}^N, \{\mathcal{T}_{T_{new}}^i\}_{i=0}^N, \{\mathcal{U}_{T_{new}}^i\}_{i=0}^N, i_{max}$;

Compute the family of $T_{new}-$steps state ahead controllable sets $\{\Xi_{T_{new}}^i\}_{i=0}^N$ via recursions (16)
Determine $\{\mathcal{T}_{T_{new}}^i\}_{i=0}^N$, $\{\mathcal{U}_{T_{new}}^i\}_{i=0}^N$ via the projections (17);
Compute the maximum index $i_{max}$ complying with the set-containment (21);

——— **Control Center** ———
(on-line)

**Input:** $\tilde{x}(t)$, $\{\Xi_{T_{new}}^i\}_{i=0}^N, \{\mathcal{T}_{T_{new}}^i\}_{i=0}^N$, $\{\mathcal{U}_{T_{new}}^i\}_{i=0}^N$, $i_{max}$, $\mathcal{J}$, $\mathcal{F}^0$;
**Output:** $u^c(t)$
**Initialization:** status=no attack, timer=0;
**Feasibility start condition:** $x(0) \in \bigcup_{i=0}^{N_{max}}\{\mathcal{T}_{T_{new}}^i\}$

**if** status== no attack **then**
  **if Detector**$(\tilde{x}(t))$==attack **then**
    status=attack
    **Disconnect** the communication medium;
    Goto 18
  **else**
    Find $i(t) = \arg\min_i : \tilde{x}(t) \in \mathcal{T}_{T_{new}}^i$
    Uniform randomly choose $j(t) \in \mathcal{J}$;
    **if** $i(t) == 0$ **then** $u^c(t) = f_{j(t)}^0(\tilde{x}(t))$
    **else**
      Compute $u^c(t)$ by (18) with $J_{j(t)}(\tilde{x}(t), u)$;
    **end if**
  **end if**
  Send $u^c(t)$ to the actuator;
**else** ▷ status=attack
  **Disconnect** the communication medium;
  **if** timer$< T_{new}$ **then**
    timer=timer+1; ▷ Channel re-connection phase
    **if** connection-reestablished==true **then**
      status==no attack; timer=0; and Goto 7;
    **end if**
  **else**
    status==no attack; timer=0; Goto 7;

24: **end if**
25: **end if**
26: $t \leftarrow t+1$, goto Step 1

---

*Theorem 1: Let the non-empty sequences* $\{\Xi^i_{T_{new}}\}_{i=0}^{N}, \{\mathcal{T}^i_{T_{new}}\}_{i=0}^{N}$, $\{\mathcal{U}^i_{T_{new}}\}_{i=0}^{N}$ *be given and* $x(0) \in \bigcup_{i=0}^{N_{max}}\{\mathcal{T}^i_{T_{new}}\}$. *Then, the supervisor logic, the actuator logic and the* $\tau$***-ST-RHC*** *algorithm ensures the constraints satisfaction and UUB property of the regulated state trajectory regardless of any admissible attack scenario.*

*Proof* - First, under free-attack scenarios, the optimization (18) is always feasible despite any disturbance realization. In fact, by construction, the $\tau-$ step controllable regions (16) guarantee that at each time instant there exists an admissible command input $u(t)$ compatible with (2). Moreover, the regulated state trajectory $x(\cdot)$ will be driven to $\mathcal{T}^0_{new}$ in at most $N$ steps and there confined in virtue of its positively invariant property. As long as DoS occurrences or FDI attacks are concerned, two scenarios come out: *1)* the controller-to-actuator channel *is not affected*; *2)* the controller-to-actuator channel *is affected*. Under *1)*, the overall feasibility is kept because, as shown in *Lemma* 1, the $T_{new}-$step controllable set guarantees that the last computed command $u^c(t) \in \mathcal{U}$ can be repeatedly applied for $T_{new}$ steps and the resulting state trajectory will be confined into $\mathcal{T}^{i(t)}_{T_{new}}$. Conversely, under *2)*, it has been shown in Section V-B that any attack event can be promptly detected as soon as the corrupted state trajectory $x(t)$ does not belong to the expected level set, i.e. $\tilde{x}(t) \notin \mathcal{T}^{i(t)}_{T_{new}}$. By denoting with $\bar{t}$ the detection time instant, a zero-input command is then repeatedly applied within the time interval $[\bar{t}, \bar{t}+T_{new}-1]$, see (21), while a guaranteed attack-free communication is restored. Finally, by noticing that the state trajectory is UUB in $\bigcup_{i=0}^{N}\{\mathcal{T}^i_{T_{new}}\}$ despite the operating scenario, the post attack recovery is always viable.

## VI. Illustrative Example

Consider the following continuous-time model [39]

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 0.8 & 0.5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t) + \begin{bmatrix} 1 \\ 1 \end{bmatrix} d(t)$$

subject to $|u(t)| \leq 5, |x_1(t)| \leq 2.5, |x_2(t)| \leq 10, |d(t)| \leq 0.05$ and discretized by means of the forward Euler method with a sampling time $T_s = 0.02\,sec$. According to *Assumptions* 1-3, a reliable communication medium with $T_{new} = 4$ time steps $(0.08sec)$ and $T_{viol} = 5$ time steps $(0.1sec)$ is exploited. The initial state condition has been chosen as $x(0) = [-1.09, 5.11]^T$ and a polyhedral family of 60 $T_{new}-$steps controllable sets has been computed, see Fig. 5, with $x(0) \in \mathcal{T}^{45}_{T_{new}}$ and the maximum set level $i_{max} = 45$.
To appreciate the *modus operandi* of the proposed supervisor-based control architecture of Fig. 2, the following attack occurrences have been considered:

- (*Attack 1: DoS on the actuation channel*) Starting from $t = 0.14\,sec$, the actuator does not receive any new packet. According to Step 2 of the **Actuator Logic**, the stored command $u(t) = u^c(0.12) = 4.95$ is safely applied because both Pre-Check and Post-Check conditions are satisfied. At

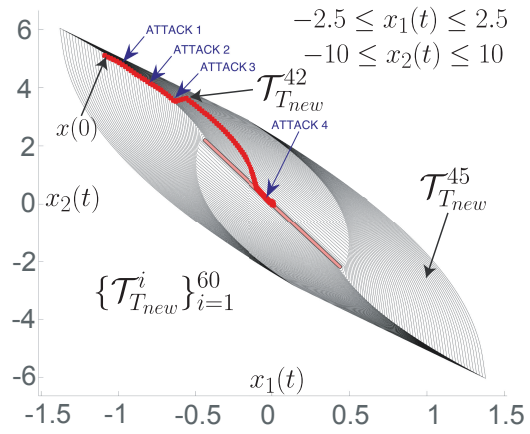


Fig. 5. $\left\{\mathcal{T}^i_{T_{new}}\right\}_{i=0}^{60}$ family (black polyhedra) and state trajectory (red solid line). Blue arrows point to the current system state vector at the beginning of each attack scenario.

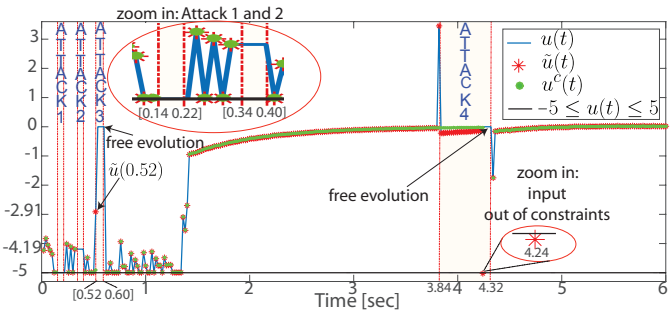


Fig. 6. Command inputs: actuator output $u(t)$, corrupted signal $\tilde{u}(t)$, computed command $u^c(t)$.

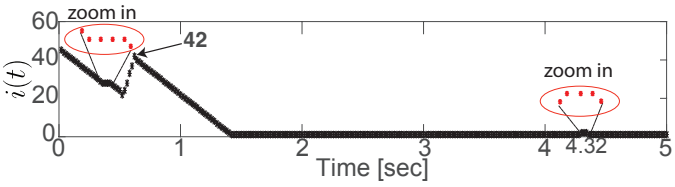


Fig. 7. Set-membership signal

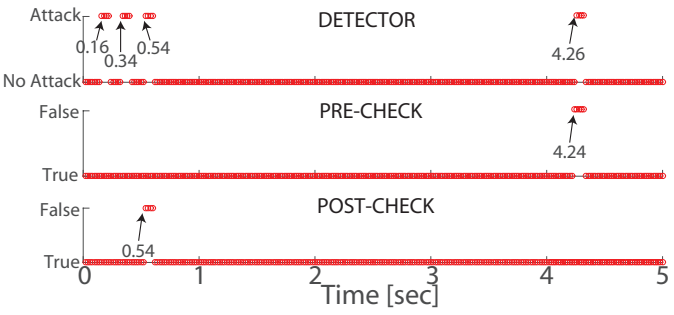


Fig. 8. Detector, Pre-Check and Post-Check flag signals.

This article has been accepted for publication in IEEE Transactions on Control of Network Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3182098



$t = 0.16sec$, the Detector identifies the attack (see Fig. 8) because $\tilde{x}(0.16) \notin \mathcal{X}^+$ where

$$\mathcal{X}^+ = \{x^+ \in \mathrm{R}^n : x^+ = A\tilde{x}(0.14) + Bu^c(0.14) + B_d d,\ d \in \mathcal{D}\}$$

As prescribed in Steps 17-24 of the $\tau$**-ST-RHC** algorithm, the existing communications are interrupted and the procedure to reestablish safe channels ends at $t = 0.24\,sec$.

- (*Attack 2: DoS on the sensor-to-controller channel*) Within the time interval $[0.34\ \ 0.40]\,sec$, the state measurements are not received by the control center, therefore the detector logic triggers an anomaly. As a consequence, the communication is disconnected and data losses occur on the actuator side: again the admissible stored command is exploited, i.e. $u(t) = u^c(0.32) = -4.19$ is applied, see Fig. 6.

- (*Attack 3: simple FDI on the actuation channel*) At $t = 0.52\,sec$, a malicious agent injects $u^a(0.52) = 2$ on the current input $u^c(0.52) = -4.91$ as prescribed in (6). Although the received command $\tilde{u}(0.52) = -2.91$ is still admissible as testified by the Pre-Check unit, the consequence is that $x(0.54) \in \mathcal{T}^{20}_{T_{new}}$ while the expected set-membership prescribes that it should belong to $\mathcal{T}^{18}_{T_{new}}$. Then, the Post-Check identities an attack and, at the next time instant, communications are blocked. From $0.53\,sec$ onward, the actuator logic set the open-loop mode, i.e. $u(t) = 0$. Although during the channel re-connection phase, which ends at $t = 0.60\,sec$, the set-membership signal $i(t)$ increases (see Fig.7), this does not compromise the feasibility retention because $x(0.53) \in \{\mathcal{T}^i_{T_{new}}\}_{i=0}^{i_{max}}$ and the zero-input state evolution will remain there confined.

- (*Attack 4: Constrained coordinated FDI attack*) At $t = 3.84$ sec, with $x(3.83) \in \mathcal{T}^0_{T_{new}}$ the coordinated attack (12)-(13) starts. The malicious control action is computed by solving the following optimization problem

$$\tilde{u}(t) = \arg\max_u ||Ax + Bu||, \quad s.t.\ Ax + Bu \in \tilde{\mathcal{T}}^0,\ \ u \in \mathcal{U}$$

aiming to keep the regulated state trajectory as far as possible from zero while remaining in the terminal region (to avoid detection), see Figs. 5, 7. The malicious agent is capable to remain stealthy until $t = 4.24\,sec$ because it is not possible to discriminate between attack occurrences and disturbance/noise realizations, i.e. $\tilde{x}(t) \in \mathcal{X}^+, \forall t \in [3.84,\ 4.24]\,sec$. In fact the corrupted command is such that $\tilde{u}(4.24) = -5.025 \notin \mathcal{U}^0_{T_{new}}$ and, as a consequence, the Pre-Check is capable to detect the anomalous event. This comes out thanks to the time-varying nature of the optimization (18) that is unknown to the malicious agent which in turn imposes $\tilde{u}(4.24) = \hat{u}^c(4.24) + u^a(4.24)$, with $u^a(4.24) = 4.993$ and $\hat{u}^c(4.24)$ the assumed value of $u^c(4.24)$. Since the malicious agent is not aware about the current performance index $J_{j(t)}(\tilde{x}(t), u)$, then it cannot exactly determine $u^c(4.24)$, but it can only guess its value. In the specific, at $t = 4.24\,\mathrm{sec}$, the attacker obtains $\hat{u}^c(4.24) = -0.032$ instead of $u^c(4.24) = -0.059$, causing $\tilde{u}(4.24) \notin \mathcal{U}^0_{T_{new}}$.
As expected, under attack-free scenario, from $t = 4.32sec$ onward, the regulated state trajectory asymptotically converges to the origin.

Finally, to highlight the capabilities of the proposed watermarked detector, in Figs. 9-10, we compare the state and control signal evolution in the presence of the basic set-based passive detector (labeled as "$J$ fixed" in the figures) and enhanced active anomaly detector (labeled as "$J$ random" in the figures). By focusing the attention on *attack 4* (highlighted regions in Figs. 9-10), it is clear that the basic passive detector is not capable to detect the attack occurrence: therefore, differently from the enhanced active detection strategy (continuous lines), the attack is never removed, and the resulting chattering-like detrimental behaviors give rise to a significant loss of performance.

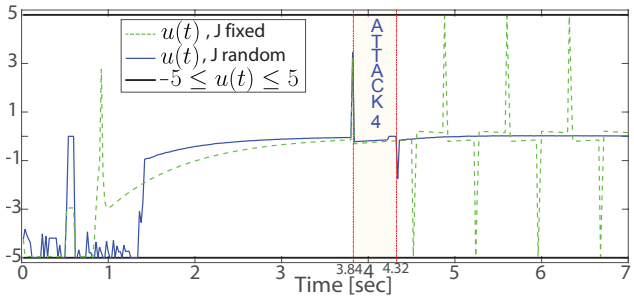


Fig. 9. Actuator output signals, $u(t)$, comparison: optimization with fixed cost function (green dotted line), optimization with random cost function (blue solid line).

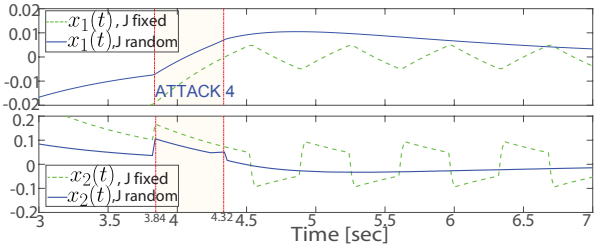


Fig. 10. Plant state vector, $x(t)$, comparison: optimization with fixed cost function (green dotted line), optimization with random cost function (blue solid line).

For the interested reader, then simulation demos are available at the following web link: https://goo.gl/wruW1T.

## VII. Conclusions

In this paper, a novel resilient control framework has been developed in order to mitigate the undesired effects arising when stealthy attacks affect the nominal behavior of constrained CPS. The idea is to combine into a unique framework the set-theoretic based RHC approach with watermarking-like arguments: this allows one to provide formal countermeasures to DoS and FDI attacks. Moreover, it is proved that constraints satisfaction and Uniform Ultimate Boundedness of the regulated system are satisfied regardless of any admissible attack occurrence. Future studies will be devoted to analyze the capability of the proposed framework to efficiently deal with stochastic scenarios by exploiting backward and forward stochastic reachability arguments in place of the robust counterparts used in this paper.

## References

[1] T. Samad and A. Annaswamy, "The impact of control technology," *IEEE Control Systems Society*, vol. 1, p. 246, 2011.

[2] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 106–117, 2016.

[3] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 4–13, 2016.

[4] M. Pajic, I. Lee, and G. J. Pappas, "Attack-resilient state estimation for noisy dynamical systems," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 82–92, 2016.

[5] R. M. Ferrari and A. M. Teixeira, "Detection and isolation of routing attacks through sensor watermarking," in *American Control Conference (ACC)*. IEEE, 2017, pp. 5436–5442.

[6] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[7] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of cps security," *Annual Reviews in Control*, vol. 47, pp. 394–411, 2019.

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

[9] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 110–127, 2015.

[10] Y. Chen, S. Kar, and J. M. Moura, "Dynamic attack detection in cyber-physical systems with side initial state information," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4618–4624, 2016.

[11] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, 2009, pp. 911–918.

[12] P. Griffioen, S. Weerakkody, and B. Sinopoli, "A moving target defense for securing cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 66, no. 5, pp. 2016–2031, 2020.

[13] C. Schellenberger and P. Zhang, "Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 1374–1379.

[14] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 82–92, 2015.

[15] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *HotSec*, 2008.

[16] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 636–653, 2009.

[17] W. M. H. Heemels, A. R. Teel, N. Van de Wouw, and D. Nešić, "Networked control systems with communication constraints: Tradeoffs between transmission intervals, delays and performance," *IEEE Transactions on Automatic control*, vol. 55, no. 8, pp. 1781–1796, 2010.

[18] S. Amin, A. A. Cárdenas, and S. S. Sastry, "Safe and secure networked control systems under denial-of-service attacks," in *International Workshop on Hybrid Systems: Computation and Control*. Springer, 2009, pp. 31–45.

[19] Y. Yuan, Q. Zhu, F. Sun, Q. Wang, and T. Başar, "Resilient control of cyber-physical systems against denial-of-service attacks," in *International Symposium on Resilient Control Systems (ISRCS)*. IEEE, 2013, pp. 54–59.

[20] A. Gupta, C. Langbort, and T. Başar, "Optimal control in the presence of an intelligent jammer with limited actions," in *IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 1096–1101.

[21] A.-Y. Lu and G.-H. Yang, "Input-to-state stabilizing control for cyber-physical systems with multiple transmission channels under denial of service," *IEEE Transactions on Automatic Control*, vol. 63, no. 6, pp. 1813–1820, 2017.

[22] M. Zhu and S. Martinez, "On the performance analysis of resilient networked control systems under replay attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 804–808, 2013.

[23] Y. Shoukry, J. Araujo, P. Tabuada, M. Srivastava, and K. H. Johansson, "Minimax control for cyber-physical systems under network packet scheduling attacks," in *ACM international conference on High confidence networked systems*, 2013, pp. 93–100.

[24] Z.-H. Pang and G.-P. Liu, "Design and implementation of secure networked predictive control systems under deception attacks," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 5, pp. 1334–1342, 2011.

[25] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "A hybrid stochastic game for secure control of cyber-physical systems," *Automatica*, vol. 93, pp. 55–63, 2018.

[26] X. Jin, W. M. Haddad, and T. Yucelen, "An adaptive control architecture for mitigating sensor and actuator attacks in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 6058–6064, 2017.

[27] L. An and G.-H. Yang, "Lq secure control for cyber-physical systems against sparse sensor and actuator attacks," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 2, pp. 833–841, 2018.

[28] F. Blanchini and S. Miani, *Set-theoretic methods in control*. Springer, 2008.

[29] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

[30] E. Fridman and U. Shaked, "Delay-dependent $h_\infty$ control of uncertain discrete delay systems," *European Journal of Control*, vol. 11, no. 1, pp. 29–37, 2005.

[31] D. Angeli, A. Casavola, G. Franzè, and E. Mosca, "An ellipsoidal off-line mpc scheme for uncertain polytopic discrete-time systems," *Automatica*, vol. 44, no. 12, pp. 3113–3119, 2008.

[32] D. P. Bertsekas and I. B. Rhodes, "On the minimax reachability of target sets and target tubes," *Automatica*, vol. 7, no. 2, pp. 233–247, 1971.

[33] G. Franzè, F. Tedesco, and W. Lucia, "Resilient control for cyber-physical systems subject to replay attacks," *IEEE Control Systems Letters*, vol. 3, no. 4, pp. 984–989, 2019.

[34] K. Ding, Y. Li, D. E. Quevedo, S. Dey, and L. Shi, "A multi-channel transmission schedule for remote state estimation under dos attacks," *Automatica*, vol. 78, pp. 194–201, 2017.

[35] O. D. Incel, "A survey on multi-channel communication in wireless sensor networks," *Computer Networks*, vol. 55, no. 13, pp. 3081–3099, 2011.

[36] N. Saputro, K. Akkaya, and S. Uludag, "A survey of routing protocols for smart grid communications," *Computer Networks*, vol. 56, no. 11, pp. 2742–2771, 2012.

[37] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[38] G. Franzè, F. Tedesco, and D. Famularo, "Model predictive control for constrained networked systems subject to data losses," *Automatica*, vol. 54, pp. 272–278, 2015.

[39] F. Blanchini and S. Miani, "Any domain of attraction for a linear constrained system is a tracking domain of attraction," *SIAM Journal on Control and Optimization*, vol. 38, no. 3, pp. 971–994, 2000.


**Walter Lucia** is an Associate Professor at the Concordia Institute for Information Systems Engineering, Concordia University, Canada. He received the M.Sc. degree in automation engineering and the Ph.D. degree in systems and computer engineering from the University of Calabria, Italy, in 2011 and 2015, respectively. In 2013, he was a Visiting Research Scholar with the ECE Department at Northeastern University, USA, and in 2015, Visiting Post-Doctoral Researcher with the ECE Department at Carnegie Mellon University, USA. His research interests include control of unmanned vehicles, fault-tolerant control, model predictive control, and resilient control of cyber-physical systems. Dr. Lucia is currently an Associate Editor of the Control System Society - Conference Editorial Board, and IEEE Systems Journal.

This article has been accepted for publication in IEEE Transactions on Control of Network Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TCNS.2022.3182098



**Giuseppe Franzè** received the Laurea degree in computer engineering in 1994 and the Ph.D. degree in systems engineering in 1999 from the University of Calabria, Italy. Since 2022 he is a Full Professor at the University of Calabria with the DIMEG department. He authored or co-authored of more than 180 research papers in archival journals, book chapters and international conference proceedings. His current research interests include constrained predictive control, nonlinear systems, networked control systems, control under constraints and control reconfiguration for fault tolerant systems, resilient control for cyber-physical systems. In November-December 2019, he was a visiting professor at Concordia University (CA) with the CIISE Department. Since 2019 he is Senior Member of IEEE. He is a co-recipient of the Best Paper Award at the IEEE-CoDIT 2019 Conference, Paris, France. He currently serves as a Associate Editor of the IEEE/CAA Journal of Automatica Sinica (JAS). He is the Guest Editor of the Special Issue Resilient Control in Large-Scale Networked Cyber-Physical Systems IEEE/CAA Journal of Automatica Sinica (JAS), 2020. From January 2018 to March 2022, he was the Graduate Program Director of the Master Degree in Automation Engineering at the DIMES department, University of Calabria.

**Bruno Sinopoli** is the Das Family Distinguished Professor at Washington University in St. Louis, where he is also the founding director of the center for Trustworthy AI in Cyber-Physical Systems and chair of the Electrical and Systems Engineering Department. He received the Dr. Eng. degree from the University of Padova in 1998 and his M.S. and Ph.D. in Electrical Engineering from the University of California at Berkeley, in 2003 and 2005 respectively. After a postdoctoral position at Stanford University, Dr. Sinopoli was member of the faculty at Carnegie Mellon University from 2007 to 2019, where he was a professor in the Department of Electrical and Computer Engineering with courtesy appointments in Mechanical Engineering and in the Robotics Institute and co-director of the Smart Infrastructure Institute. His research interests include modeling, analysis and design of Resilient Cyber-Physical Systems with applications to Smart Interdependent Infrastructures Systems, such as Energy and Transportation, Internet of Things and control of computing systems.