A Zonotopic-Based Watermarking Design to Detect Replay Attacks

Carlos Trapiello and Vicenç Puig

***Abstract*—This paper suggests the use of zonotopes for the design of watermark signals. The proposed approach exploits the recent analogy found between stochastic and zonotopic-based estimators to propose a deterministic counterpart to current approaches that study the replay attack in the context of stationary Gaussian processes. In this regard, the zonotopic analogous case where the control loop is closed based on the estimates of a zonotopic Kalman filter (ZKF) is analyzed. This formulation allows to propose a new performance metric that is related to the Frobenius norm of the prediction zonotope. Hence, the steady-state operation of the system can be related with the size of the minimal Robust Positive Invariant set of the estimation error. Furthermore, analogous expressions concerning the impact that a zonotopic/Gaussian watermark signal has on the system operation are derived. Finally, a novel zonotopically bounded watermark signal that ensures the attack detection by causing the residual vector to exit the healthy residual set during the replay phase of the attack is introduced. The proposed approach is illustrated in simulation using a quadruple-tank process.**

***Index Terms*— Optimal control, physical watermarking, replay attack, zonotopes.**

I. Introduction

DURING the last decade, the interest aroused by cybersecurity aspects of networked control systems (NCSs) has experienced an exponential growth, currently representing one of the main trends in NCSs [1]. Some of the factors that support this increasing interest are: the potential risks to human lives as well as significant economic losses as a consequence of successful attacks; an ever-growing list of registered attacks, including famous cases like the Stuxnet malware [2]; the identification of possible threats and vulnerabilities on NCSs [3]–[5]; the characterization stealthy attack policies which exploit the limitations of traditional monitoring algorithms [6], etc. Consequently, all of the above has motivated an exhaustive analysis on secure control and state estimation techniques for the case where a malicious adversary is capable of attacking either the physical components of the system or the communication network that is used to close the control loop [7], [8].

In particular, in the present paper the focus is on the detection of replay attacks from a model-based perspective (i.e., testing the compatibility of measurements with the model of the underlying physical process), thus working under the premise that classical information security detection techniques, like timestamp, can be bypassed by an attacker. On this subject, note that an effective detection is required for deploying replay attack resilient control schemes [9], [10]. Otherwise, regarding model-based detection techniques: in the pioneering work [11], the authors not only model and analyze the effect of replay attacks in the framework of stationary Gaussian processes, but also propose the injection of an exogenous signal into the control loop in order to improve the detectability of the attack. From this moment on, the design of these exogenous signals, which are denoted with the term *physical watermarking* [12], has been intensively studied until becoming a well-adopted technique for the detection of replay attacks [13].

Related Work: The watermarking approaches that have been proposed in the automatic control literature can be broadly classified depending on whether they are injected in the control loop in an additive or multiplicative way. On the one hand, multiplicative watermarking schemes employ a bank of filters, placed on the plant side of the communications network, in order to watermark the sensor outputs, while the original data are reconstructed by a watermarking remover placed on the controller side [14], [15]. These schemes allow to address the closed-loop performance degradation induced by the watermark signal. However, multiplicative schemes may lose their effectiveness if the attacker is able to replace the real data before the addition of the watermark signal.

On the other hand, in additive watermarking schemes an exogenous signal is injected from the controller side at the cost of inducing some closed-loop performance degradation. In this regard, in [11], [12], an additive Gaussian signal is injected in the control loop for increasing the attack detection rate of a statistical detector. A similar watermarking scheme has been also analyzed in [16]. In [17], a Gaussian signal is injected in a combined algorithm that deals with online watermarking design and unknown system parameters identification. Furthermore, in [18], a periodical injection of a Gaussian signal is proposed for detecting discontinuous replay

Manuscript received December 22, 2021; revised January 28, 2022; accepted February 16, 2022. This work was in part supported by the Margarita Salas grant from the Spanish Ministry of Universities funded by the European Union NexGenerationEU, and in part co-funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERFD) through the project SaCoAV (ref. MINECO PID2020-114244RB-I00). Recommended by Associate Editor Xin Luo. *(Corresponding author: Carlos Trapiello.)*

Citation: C. Trapiello and V. Puig, " A zonotopic-based watermarking design to detect replay attacks," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 11, pp. 1924–1938, Nov. 2022.

C. Trapiello was with the Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politécnica de Catalunya (UPC), Rambla Sant Nebridi, 22, 08222 Terrassa, Spain. He is now with the UMR 5218 - IMS Laboratory, France (e-mail: carlos.trapiello@upc.edu).

V. Puig is with the Supervision, Safety and Automatic Control Research Center (CS2AC) of the Universitat Politécnica de Catalunya (UPC), Rambla Sant Nebridi, 22, 08222 Terrassa, and also with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Carrer Llorens Artigas, 4-6, 08028 Barcelona, Spain (e-mail: vicenc.puig@upc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.105944

attacks. A stochastic game approach is proposed in [19] in order to derive an optimal control policy that switches between a control-cost optimal (but nonsecure) and watermarked (but cost-suboptimal) controllers. Moreover, in [20], a dynamic watermarking method is proposed for detecting replay attacks in linear-quadratic Gaussian (LQG) systems.

At this point, it must be highlighted that the additive schemes presented above have addressed the physical watermarking design problem from a stochastic point of view and, in particular, within the context of stationary Gaussian processes. Nevertheless, as remarked in [21], despite the characterization of uncertainties through Gaussian probability distributions is often well suited to deal with measurement noises, this description may fail to model disturbances that arise from some of lack of knowledge about deterministic behaviors and which may do not have any other stationary behavior than that of remaining within some specified bounds. Hence, in the remaining of the paper the physical watermarking design problem will be addressed from a set-based (or norm-bounded) paradigm [22]. These set-based techniques have proven their effectiveness in fault diagnosis and tolerant control schemes [23]. Besides, they have been recently used in order to achieve a secure state estimation in systems under bounded attacks [24], [25] and to characterize the replay attack detectability using set invariance notions [26].

Approaching the system monitoring under a set-based paradigm allows to infer in a deterministic way if the system is being attacked by testing online whether or not a residual signal belongs to a set coherent with the healthy operation of the system. Nonetheless, this unambiguous assessment poses two central challenges in the design of active attack detection schemes: i) The proposed watermarking scheme should be able to force the residual signal out of its healthy residual set whenever the system is under attack while minimizing the performance loss induced in the system operation; ii) A coherent metric to assess the performance degradation induced by a bounded watermark signal should be developed.

In order to address the latter point, zonotopic sets will be used mainly motivated by the results presented in [27], where the notion of covariation matrix of a zonotope is introduced for proposing an optimal robust state observer. Furthermore, by minimizing the weighted Frobenius norm (F-norm) of the prediction zonotope, analogous expressions are obtained regarding the optimal design of a zonotope-based observer and the standard stochastic Kalman filter (uncertainties modeled as Gaussian random variables). Consequently, in this paper, the above analogy is further exploited by analyzing the optimal control law that minimizes a cost function that takes into account the weighted F-norm of the bounding zonotope computed by a zonotopic Kalman filter (ZKF).

It must be pointed out that optimal control has been thoroughly studied under different uncertainty paradigms, namely: *deterministic* optimal control [28], where uncertainty can only take one value; *stochastic* optimal control [28]–[31], where uncertainty is defined as a function of a probability space; *minimax* control [32]–[34], where uncertainty is confined within a given set. In this regard, notice that, whereas minimax control is defined in a worst case basis, the proposed approach is based on the bounding zonotope F-norm which, for linear-quadratic control problems, returns expressions similar to those obtained if the problem is formulated under Gaussian uncertainties. Consequently, when it comes to the active attack detection in linear systems, through this novel framework the same optimality patterns are obtained regardless of whether the detection is approached under a set-based or stochastic paradigm, what may lead to future detection schemes that merge the mutually exclusive benefits of both approaches [21].

Contributions: The contributions of this paper are as follows. First, a linear-quadratic regulator (LQR) controller that operates based on the estimates of a ZKF is proposed. It is proven that this control scheme minimizes a new cost function that considers the weighted F-norm of the system states, in what can be seen as the zonotopic counterpart of the expected value of a quadratic cost function in an LQG scheme. Furthermore, for the infinite horizon control problem, the Frobenius radius of the minimum robust positive invariant set (mRPI) of the estimation error is given by the solution of the Riccati algebraic equation.

Second, the new zonotope-based cost function is used to evaluate the impact that a zonotopically bounded watermark signal has on the steady-state operation of the system. Analogous expressions to the performance loss induced by an exogenous Gaussian signal in an LQG system are obtained.

Third and finally, a zonotopically bounded watermark signal that minimizes the new cost function is proposed. This detection scheme injects a known signal in the system inputs, while filtering its effect through the estimations generated using the output data. Since the replay phase of the attack entails that the known signal is no longer observable, then the estimation error generates a new residual signal that is destabilized whenever the output data are being replayed back. Some preliminary results have been presented in the conference paper [35].

The remainder of the paper is organized as follows: Section II introduces some preliminary concepts whereas Section III presents the problem statement. Then, in Section IV the optimal finite/infinite horizon control problem based on the estimates of a ZKF is analyzed. Section V analyzes the impact of a bounded watermark signal on the previous optimal control loop. Besides, in Section VI a new guaranteed replay attack detection scheme is introduced while in Section VII the proposed results are validated in simulation using a well-known control benchmark based on a four-tank system. Finally, in Section VIII the main conclusions are drawn.

II. Notation, Basic Definitions and Properties

A. Notation

The following notations are used along this work. $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the n and $m\times n$ dimensional Euclidean space, respectively. $\mathbb{N}$ denotes the set of non-negative integers. The identity matrix of dimension n is denoted by I_n and 0 denotes a null matrix whose dimensions can be deduced from the context. In the sequel, the symbols $\geq, >, \leq, <$ and $|\cdot|$ should be understood elementwise. The trace of a square matrix X is denoted by $Tr[X] = \sum_i X_{ii}$. For a matrix X, $X > 0$ ($X \geq 0$) means that X is a positive definite (semi-definite) matrix.

Besides, given the matrices A, B, C and the vectors x, y of appropriate dimensions, then the following identities are satisfied:

$$Tr[A] = Tr[A^T], \qquad Tr[Ayx^T] = x^T Ay,$$
$$Tr[ABC] = Tr[BCA] = Tr[CAB]. \tag{1}$$

B. Basic Definitions and Properties

The Minkowski sum of two sets $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined by $\mathcal{S}_1 \oplus \mathcal{S}_2 = \{s_1 + s_2 : s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2\}$ and the linear image of the set $\mathcal{S} \subset \mathbb{R}^n$ by $L \in \mathbb{R}^{q \times n}$ is $L\mathcal{S} = \{Ls, s \in \mathcal{S}\}$. Furthermore, let us introduce a discrete-time linear time-invariant (LTI) system of the form

$$x_{k+1} = Ax_k + Bw_k \tag{2}$$

where $x_k \in \mathbb{R}^n$ is the state vector at the sampling instant $k \in \mathbb{N}$, and $w_k \in \mathbb{R}^{n_w}$ is an unknown disturbance contained in the convex and compact set $\mathcal{W} \subset \mathbb{R}^{n_w}$ that contains the origin. Additionally, A and B are matrices with adequate dimensions, and A is assumed to be an asymptotically stable matrix, i.e., all its eigenvalues are strictly inside the unit disk.

Definition 1: The set $\Omega \subset \mathbb{R}^n$ is said to be *robustly positively invariant* (RPI) for the system (2), if for all $x_0 \in \Omega$ and all $w_k \in \mathcal{W}$ the solution is such that $x_k \in \Omega$ for $k > 0$. Equivalently, Ω is RPI if and only if $A\Omega \oplus B\mathcal{W} \subseteq \Omega$.

Definition 2: The *minimal RPI* (mRPI) set of (2) is the RPI set in $\mathbb{R}^n$ that is contained in every closed RPI set of (2).

The mRPI set exists, is unique, compact and contains the origin of the state-space if $\mathcal{W}$ contains the origin [36, Sec. IV]. Furthermore, from the linearity and asymptotic stability of (2), it follows that the mRPI set is the limit set of all trajectories of system (2). For a comprehensive analysis on set invariance the reader is referred to [36], [37].

Definition 3: A *zonotope* $\langle c, H \rangle \subset \mathbb{R}^n$ is the affine transformation of a unitary hypercube $\mathbf{B}^m = [-1, 1]^m$:

$$\langle c, H \rangle = c \oplus H\mathbf{B}^m = \{c + Hz : z \in \mathbf{B}^m\}$$

where $c \in \mathbb{R}^n$ is the center and $H \in \mathbb{R}^{n \times m}$ the generators matrix. The order of $\langle c, H \rangle$ is given by m/n.

Property 1 (Zonotopic operations [27]): For zonotopes, the following properties hold:

$$\langle p_1, H_1 \rangle \oplus \langle p_2, H_2 \rangle = \langle p_1 + p_2, [H_1, H_2] \rangle \tag{3a}$$

$$L\langle p, H \rangle = \langle Lp, LH \rangle \tag{3b}$$

$$\langle p, H \rangle \subseteq \langle p, b(H) \rangle \tag{3c}$$

where $p, p_1, p_2 \in \mathbb{R}^n$, $H \in \mathbb{R}^{n \times m}$, $H_1 \in \mathbb{R}^{n \times m_1}$, $H_2 \in \mathbb{R}^{n \times m_2}$ and $L \in \mathbb{R}^{l \times n}$. Besides, $\langle p, b(H) \rangle$ denotes the bounding box where $b(H) \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose entries are given by $b(H)_{ii} = \sum_{j=1}^{m} |H_{ij}|$.

Property 2 (Reduction operator $\downarrow_{q,W}$ [38]): Given the zonotope $\langle c, H \rangle \subset \mathbb{R}^n$ with $c \in \mathbb{R}^n$ and $H \in \mathbb{R}^{n \times m}$, sort the columns of H on decreasing weighted norm $\|\cdot\|_W$, that is, $H = [h_1, \ldots, h_j, \ldots, h_m]$ with $\|h_j\|_W^2 \geq \|h_{j+1}\|_W^2$. Then, by computing $\downarrow_{q,W}(H) = [H^a, b(H^b)] \in \mathbb{R}^{n \times q}$ with $H^a = [h_1, \ldots, h_{q-n}]$ and $H^b = [h_{q-n+1}, \ldots, h_m]$, it follows that $\langle c, H \rangle \subseteq \langle c, \downarrow_{q,W}(H) \rangle$.

In addition, in the present paper the size of a given zonotope is assessed by means of its F_W-radius defined below.

Definition 4 (F_W-radius [27]): Let $W \in \mathbb{R}^{n \times n}$ be a symmetric positive definite (SPD) matrix. The weighted Frobenius radius (F_W-radius) of the zonotope $\langle c, H \rangle \subset \mathbb{R}^n$ is the weighted F-norm of H, i.e., $\|\langle c, H \rangle\|_{F,W} = \|H\|_{F,W} = \sqrt{Tr[H^T W H]}$.

The F-radius of a zonotope equals the F_W-radius for $W = I_n$. Note that, since $\|H\|_{F,W}^2 = \sum_{i=1}^{m} \|h_i\|_W^2$, the F_W-radius size criterion involves the norm of all the generators of the zonotope. The Frobenius radius is a standard criterion for assessing the size of a zonotope [39]–[42].

Definition 5 (Covariation matrix [27]): The covariation matrix of the zonotope $\mathcal{Z} = \langle c, H \rangle$ is $P = HH^T$.

Hence, minimizing the F-radius of a zonotope $\langle c, H \rangle$ is equivalent to minimizing the trace of its covariation matrix, that is, $\|H\|_F^2 = Tr[P]$.

III. System Overview

The present work focuses on a perturbed system that is remotely controlled through a state estimate control policy and monitored by means of a residual generation/evaluation block. In particular, due to the unknown-but-bounded description of the uncertainties, a ZKF is used for generating the state estimate required for: i) Closing the control loop; ii) Generating the residual signal and a residual set that characterizes the healthy operation of the system. Below, the system under study and the ZKF are detailed.

A. System Description

Let us consider a discrete-time LTI perturbed system of the form

$$x_{k+1} = Ax_k + Bu_k + Ew_k \tag{4a}$$

$$y_k = Cx_k + Fv_k \tag{4b}$$

where A, B, C, E and F are the state-space matrices with adequate dimensions, $x_k \in \mathbb{R}^{n_x}$ is the state vector, $u_k \in \mathbb{R}^{n_u}$ an exogenous input signal and $y_k \in \mathbb{R}^{n_y}$ the output vector. Besides, the system state at $k = 0$ is assumed to satisfy $x_0 \in \langle c_0, H_0 \rangle$, for some $c_0 \in \mathbb{R}^{n_x}$ and $H_0 \in \mathbb{R}^{n_x \times n_0}$.

In the sequel, it is considered that for all $k \in \mathbb{N}$ process disturbances $w_k \in \mathbb{R}^{n_w}$ and sensor noise $v_k \in \mathbb{R}^{n_v}$ satisfy

$$w_k \in \langle 0, I_{n_w} \rangle, \quad v_k \in \langle 0, I_{n_v} \rangle. \tag{5}$$

Assumption 1: The pairs (A, B) and (A, C) are assumed to be stabilizable and detectable, respectively.

Remark 1: Notice that, according to Definition 3, a zonotope $\mathcal{Z} = \langle 0, H \rangle$, with $H \in \mathbb{R}^{n \times p}$, can be expressed as $\mathcal{Z} = H\langle 0, I_p \rangle$. Thus, given any zero-centered zonotope, a unitary box representation like (5) can be obtained through a coherent modification of matrices E and F.

B. Zonotopic KF

The system model (4) and the bounded uncertainties (5), are used in order to generate the zonotope-based state estimator. According to [27], given an initial state $x_0 \in \langle c_0, H_0 \rangle$, then recursively defining the center estimate c_k and the generators matrix H_k as

$$c_{k+1} = (A - G_k C)c_k + Bu_k + G_k y_k \tag{6a}$$

$$H_{k+1} = [(A - G_k C)\check{H}_k, \; E, \; -G_k F] \tag{6b}$$

$$\check{H}_k = \downarrow_{q,W}(H_k) \tag{6c}$$

the state inclusion property $x_k \in \langle c_k, H_k \rangle$ holds for all $k \geq 0$.

Moreover, the optimal observer gain that minimizes the F_W-radius of the prediction zonotope $\langle c_{k+1}, H_{k+1} \rangle$, i.e., $G_k^* = \arg\min_{G_k} \|H_{k+1}\|_{F,W}^2$, is computed as [27]

$$G_k^* = AK_k^* = A\check{P}_k C^T (C\check{P}_k C^T + Q_v)^{-1} \tag{7}$$

where $P_k, \check{P}_k, Q_w \in \mathbb{R}^{n_x \times n_x}$ and $Q_v \in \mathbb{R}^{n_y \times n_y}$ are the covariation matrices

$$P_k = H_k H_k^T, \quad \check{P}_k = \check{H}_k \check{H}_k^T, \quad Q_w = EE^T, \quad Q_v = FF^T$$

and matrix P_k satisfies

$$P_{k+1} = A\check{P}_k A^T + Q_w - A\check{P}_k C^T (C\check{P}_k C^T + Q_v)^{-1} C\check{P}_k A^T. \tag{8}$$

In the sequel, it is considered that $Q_v \succ 0$ and that the pair (A, Q_w) is stabilizable. Notice that the weighted criterion in the F_W-radius is selected in order to be consistent with the reduction operator $\downarrow_{q,W}$ presented in Property 2. In this regard, if the generators matrix is sorted as $H_k = [H_k^a, H_k^b]$ (in such a way that $\check{H}_k = \downarrow_{q,W}(H_k) = [H_k^a, b(H_k^b)]$), then the difference in the covariation matrices induced by the reduction operator can be rewritten as

$$\check{P}_k = P_k - H_k^b H_k^{bT} + b(H_k^b)^2. \tag{9}$$

IV. Linear-Quadratic Zonotopic Control

In this section, the optimal control problem for the case in which the feedback loop is closed using the estimates generated by a ZKF is analyzed. Note that, under this paradigm, at each time instant k the real state of the system is known to be confined within a zonotope. In this regard, the following performance criterion is introduced in order to assess the system operation.

Definition 6 (Zonotopic quadratic performance): Given an SPD matrix $S \in \mathbb{R}^{n \times n}$ and the unknown but zonotopically bounded vector $x \in \langle c, H \rangle \subset \mathbb{R}^n$, the performance of x is evaluated according to

$$Q[x^T S x] = c^T S c + \|H\|_{F,S}^2 = c^T S c + Tr[SP] \tag{10}$$

with $P = HH^T$.

It must be highlighted that Definition 6 matches the expected value for a Gaussian random vector centered at c and with covariance P, i.e., $x \sim \mathcal{N}(c, P)$. Furthermore, from Definition 6, it is straightforward to see that the operator $Q[\cdot]$ satisfies the following property.

Property 3 (Distributive property): Given the variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ such that $x \in \langle c_x, H_x \rangle \subset \mathbb{R}^n$ and $y \in \langle c_y, H_y \rangle \subset \mathbb{R}^m$, then

$$Q[x^T S_x x + y^T S_y y] = Q[x^T S_x x] + Q[y^T S_y y] \tag{11}$$

with S_x and S_y SPD matrices of appropriate dimensions.

Proof: The left-hand side of (11) can be rewritten as

$$Q\left[\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right] = \lambda$$

with

$$\begin{bmatrix} x \\ y \end{bmatrix} \in \left\langle \begin{bmatrix} c_x \\ c_y \end{bmatrix}, \begin{bmatrix} H_x & 0 \\ 0 & H_y \end{bmatrix} \right\rangle.$$

Therefore, from Definition 6, it follows:

$$\begin{aligned} \lambda &= \begin{bmatrix} c_x^T & c_y^T \end{bmatrix} \begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} c_x \\ c_y \end{bmatrix} \\ &\quad + Tr\left[\begin{bmatrix} S_x & 0 \\ 0 & S_y \end{bmatrix} \begin{bmatrix} H_x & 0 \\ 0 & H_y \end{bmatrix} \begin{bmatrix} H_x^T & 0 \\ 0 & H_y^T \end{bmatrix} \right] \\ &= c_x^T S_x c_x + Tr[S_x P_x] + c_y^T S_y c_y + Tr[S_y P_y] \\ &= Q[x^T S_x x] + Q[y^T S_y y]. \end{aligned}$$

■

Below, the finite and infinite horizon control problems under the performance metric in Definition 6 are analyzed.

A. Finite Horizon Control

Consider a non-measurable state x_k for which a zonotopic state estimation $\langle c_k, \check{H}_k \rangle \supseteq \langle c_k, H_k \rangle$ is generated by means of the ZKF described in Section III-B. The finite horizon linear-quadratic control is posed as computing the inputs sequence $\{u_k : k = 0, \ldots, N-1\}$ which minimizes the following cost function:

$$J(N, x_0, u) = Q\Big[\sum_{k=0}^{N-1} (x_k^T W x_k + u_k^T U u_k) + x_N^T W_f x_N \Big] \tag{12}$$

where W, W_f and U are SPD matrices, the pair (A, W) is detectable and $x_0 \in \langle c_0, H_0 \rangle$.

At a generic instant k_1 (for which the state satisfies $x_{k_1} \in \langle c_{k_1}, \check{H}_{k_1} \rangle$), then the optimal *cost-to-go* of (12) is given by the expression

$$V_{k_1}(x_{k_1}) = \min_{u_{k_1}^*, \ldots, u_{N-1}^*} \left\{ Q\Big[\sum_{k=k_1}^{N-1} (x_k^T W x_k + u_k^T U u_k) + x_N^T W_f x_N \Big] \right\}.$$

Moreover, from dynamic programming and taking into account Property 3, it is known that

$$V_k(x_k) = \min_{u_k^*} \left\{ Q[x_k^T W x_k + u_k^T U u_k] + V_{k+1}(x_{k+1}) \middle| \hat{X}_k \right\}$$

where $\hat{X}_k$ represents the set of estimations $\hat{X}_k = \{x_0, \ldots, x_k\}$, such that $x_k \in \langle c_k, \check{H}_k \rangle$, and $V_{k+1}(x_{k+1}) | \hat{X}_k$ is the *cost-to-go* obtained at $k+1$ given the sequence $\hat{X}_k$. Based on this, the following theorem is presented.

Theorem 1: At each time instant, $V_k(x_k)$ is given by

$$V_k(x_k) = Q[c_k^T S_k c_k] + s_k \tag{13}$$

with S_k and s_k defined by the backwards recursions

$$S_k = A^T S_{k+1} A + W - A^T S_{k+1} B (B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A \tag{14a}$$

$$s_k = s_{k+1} + Tr[W\check{P}_k + S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})] \tag{14b}$$

starting at $S_N = W_f$ and $s_N = Tr[W_f \check{P}_N]$.

Proof: See Appendix A. ■

Therefore, according to the development presented in Appendix A, the control law that minimizes the cost function (12) is the LQR controller

$$u_k^* = -(B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A c_k = -L_k^* c_k \tag{15}$$

which operates based on the center of the ZKF estimator.

In order differentiate (15) from the stochastic LQG control, in the sequel the zonotopic version will be denoted as linear-quadratic zonotopic (LQZ) control. For this case, from (13) and (14) it follows that the optimal value of the cost function is:

$$J(N, x_0, u) = V_0(x_0) = \mathcal{Q}[c_0^T S_0 c_0] + s_0 \tag{16}$$

with $\mathcal{Q}[c_0^T S_0 c_0] = c_0^T S_0 c_0$ and

$$s_0 = \sum_{k=0}^{N-1} \Big(Tr[W\check{P}_k + S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})] \Big) + Tr[W_f \check{P}_N].$$

At this point, it is worth noting the fact that, if no over-approximation is introduced during the horizon N (i.e., if $\check{P}_k = P_k$), then (16) is equal to the cost function obtained in an LQG system but substituting the covariation matrices for the appropriate covariance matrix of the Gaussian variables. Therefore, it can be concluded that an optimal control scheme has the same impact on the expected value of the state variables when the uncertainties are modeled as Gaussian variables, that on the F_W-radius of the state bounding zonotope under bounded uncertainties.

Remark 2: According to Theorem 11 in [27], if the zonotope $\langle 0, H_k \rangle = \langle 0, [H_k^a,\ H_k^b] \rangle$, whose columns have been sorted in decreasing weighted norm, is over-approximated by $\langle 0, \check{H}_k \rangle = \langle 0, \downarrow_{q,W} (H_k) \rangle = \langle 0, [H_k^a\ b(H_k^b)] \rangle$, then there exists a value $q_0 \in \mathbb{N}_+$ such that by selecting $q \geq q_0$ it can be ensured that the resulting sequence of zonotopes has bounded F_W-radius. Nevertheless, under the effect of the reduction operator, the gain G_k^* and covariation $\check{P}_k$ matrices may not converge to some fixed values. This is due to the coupling between: i) The convergence of the over-approximated zonotope $\langle 0, \check{H}_k \rangle$ to a fixed structure, i.e., such that $b(H_k^b)^2 - H_k^b H_k^{bT}$ becomes constant; ii) The evolution of the covariation recursion (8) which depends on the relationship $\check{P}_k - P_k = b(H_k^b)^2 - H_k^b H_k^{bT}$.

B. Infinite Horizon Control

Since many control systems operate for long periods of time, the remainder of the paper will focus on a system whose control law is obtained by solving the infinite horizon problem formulated as

$$J_\infty = \lim_{N\to\infty} \frac{1}{N} \mathcal{Q}\Big[\sum_{k=0}^{N-1} (x_k^T W x_k + u_k^T U u_k) \Big] \tag{17}$$

where X and U are SDP matrices and the state x_k is not measurable. Because the infinite horizon problem does not depend on the time to go (i.e., it is shift invariant), the following well-known time-invariant controller and observer gains are obtained:

Control: From Assumption 1, matrices W and U being SPD and the detectability of (A, W), then the optimal infinite horizon controller matches the steady-state finite horizon control given by the constant control gain

$$L^* = (B^T S_\infty B + U)^{-1} B^T S_\infty A$$

where S_∞ is the unique positive definite solution of the discrete Algebraic Riccati equation (DARE) [43]

$$S_\infty = A^T S_\infty A + W - A^T S_\infty B (B^T S_\infty B + U)^{-1} B^T S_\infty A \tag{18}$$

such that the matrix $A - BL^*$ is asymptotically stable.

Estimation: From Assumption 1, matrix Q_v being SPD and $Q_w = Q_w^T \geq 0$ and the stabilizability of the pair (A, Q_w), then the optimal infinite horizon observer gain is

$$G^* = AP_\infty C^T (CP_\infty C^T + Q_v)^{-1}$$

with P_∞ being the unique positive definite solution of the DARE

$$P_\infty = AP_\infty A^T + Q_w - AP_\infty C^T (CP_\infty C^T + Q_v)^{-1} CP_\infty A^T \tag{19}$$

such that the matrix $A - G^* C$ is asymptotically stable.

Henceforth, the estimation error is defined as $e_k = x_k - c_k$. Thus, from the comparison of (4a) and (6a) for the fixed gain G^*, it follows that the dynamics of the estimation error are governed by:

$$e_{k+1} = (A - G^* C) e_k + E w_k - G^* F v_k \tag{20}$$

in such a way that $e_{k+1} \in \langle 0, H_{k+1} \rangle$ is satisfied for

$$H_{k+1} = [(A - G^* C) H_k,\ E,\ -G^* F] \tag{21}$$

and any initial $e_0 \in \langle 0, H_0 \rangle$. Besides, the asymptotic stability of the closed-loop system (20) guarantees the convergence of its trajectories towards its mRPI set denoted as Φ. On this subject, the following theorem is introduced.

Theorem 2: The F-radius of the mRPI set Φ of system (20) is $\|\Phi\|_F^2 = Tr[P_\infty]$.

Proof: See Appendix B. ■

Regarding the computation of the mRPI set, it is well-known that its exact computation can only be achieved under the restrictive assumption that the system dynamics are nilpotent [44]. Hence, in the sequel, the set Φ will be outer-approximated through an RPI set. In this regard, by recursively refining an outer RPI set, Φ can be approximated with arbitrary precision at the price of increasing the complexity of the over-approximating set [45], [46]. The computation of an initial zonotopic RPI set is discussed in Appendix C. Accordingly, hereinafter it is considered that the system is monitored in the steady-state using a reduced order RPI zonotopic set $\langle 0, H^{ss} \rangle$ such that $\langle 0, H^{ss} \rangle \supseteq \Phi$.

For the ideal scenario where the mRPI set Φ is used in order to bound the estimation error steady-state value, the optimal infinite horizon cost function is given by

$$J_\infty = \lim_{N\to\infty} \frac{1}{N} \mathcal{Q}\Big[\sum_{k=0}^{N-1} x_k^T W x_k + u_k^T U u_k \Big] = Tr[WP_\infty + S_\infty (G^* (CP_\infty C^T + Q_v) G^{*T})]. \tag{22}$$

On the other hand, if the zonotopic over-approximation is used instead, the cost function is

$$\tilde{J}_\infty = Tr[WP^{ss} + S_\infty (G^* (CP^{ss} C^T + Q_v) G^{*T})] \tag{23}$$

with $P^{ss} = H^{ss} H^{ss,T} \neq P_\infty$.

V. Replay Attack Detectability

In this section, the case where a malicious attacker can

access the communications layer and record/replay the data sent through the sensors-to-controller link is analyzed. In order to characterize the effect of the attack, the following time windows are defined:

1) Record Window: Output data are assumed to be recorded for $K_{REC} = \{k \in \mathbb{N} : k \in [k_0,\ k_0 + l - 1]\}$, where $l \in \mathbb{N}$ denotes the size of the record window.

2) Replay Window: Real data are replaced for $K_{REP} = \{k \in \mathbb{N} : k \in [k_1,\ k_1 + l - 1]\}$.

In the sequel, whenever variables from different time windows are being compared, superscripts ^{r} and ^{a} will be used in order to differentiate that a system variable is in the record and in the replay phase, respectively. That is, for example, for the state variable x_k, for all $k \in K_{REP}$ it is denoted $x_k^a = x_k$, whereas $x_k^r = x_{k+(k_0-k_1)}$.

Remark 3: In the time windows defined above, it is implicitly assumed that the recorded sequence is only replayed once. Nevertheless, due to the linearity of the systems under study, the same analysis holds if the recorded sequence is replayed several times in a loop, just by shifting the initial time instant k_1 to match the start of each replayed sequence.

A. Anomaly Detector

The system operation is assessed based on the values adopted by the residual signal

$$r_k = y_k - Cc_k = Ce_k + Fv_k \tag{24}$$

with $e_k = x_k - c_k \in \langle 0, H_k \rangle$.

Assumption 2: The system is in stationary operation such that $k \geq k^*$, with $k^* \in \mathbb{N}$ being a finite time instant for which the estimation error satisfies $e_{k^*} \in \langle 0, H^{ss} \rangle$.

Therefore, from Assumption 2, for all $k \geq k^*$ the residual signal satisfies

$$r_k \in \mathcal{R}_H$$

where the healthy residual set $\mathcal{R}_H = \langle c_r, H_r \rangle$ is computed as

$$\langle c_r, H_r \rangle = C\langle 0, H^{ss} \rangle \oplus F\langle 0, I_{n_v} \rangle = \langle 0, [CH^{ss},\ F] \rangle.$$

Hence, the presence of anomalies is assessed according to

$$\begin{cases} r_k \in \mathcal{R}_H & \Longrightarrow \text{ Healthy operation,} \\ \text{otherwise} & \Longrightarrow \text{ Something is wrong.} \end{cases} \tag{25}$$

Remark 4: Note that the stability of (20) guarantees the convergence of its trajectories towards Φ, and thus the existence of the finite instant k^* for which Assumption 2 is satisfied. In addition, from the invariance properties of $\langle 0, H^{ss} \rangle$ it follows that $e_k \in \langle 0, H^{ss} \rangle$ for all $k \geq k^*$.

B. Attack Detectability

Fig. 1 shows the overall control/attack/watermarking scheme under consideration. Regarding the control scheme: the control loop is closed remotely by means of an optimal controller (Section IV) that operates based on the estimates of a ZKF (Section III-B). The estimates of the ZKF are used by a set-based anomaly detector (Section V-A). Additionally, Fig. 1 depicts an attacker that may be able to spoof the data sent through the sensors-to-controller link as well as to inject an exogenous signal $\bar{u}_k$ into the system inputs. On the other hand, the blue dashed box shows the interaction of an additive watermark signal (Section VI) with the control loop.

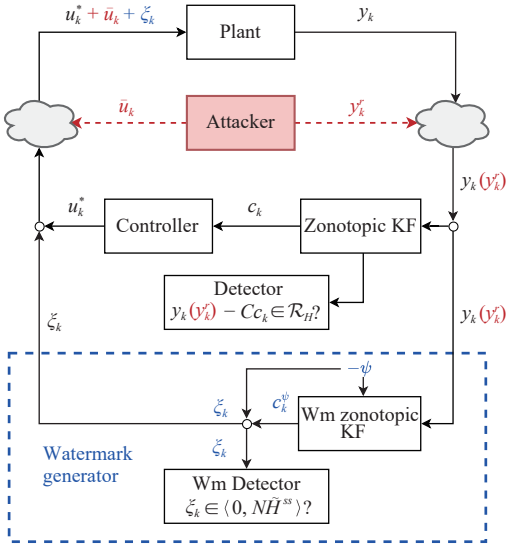


Fig. 1. Overall scheme – Replay attack (red); Watermark generator (blue).

Taking into consideration the time windows defined above, the residual signal during the replay phase of the attack is

$$r_k^a = y_k^r - Cc_k^a = y_k^r - Cc_k^r + C(c_k^r - c_k^a) = r_k^r + C(c_k^r - c_k^a) \tag{26}$$

with $r_k^r \in \mathcal{R}_H$ from Assumption 2.

Besides, the last term of (26) represents the difference between the centers of the estimates (cf. (6a)) in the record c_k^r and replay c_k^a phases. Under the optimal control law $u_k^* = -L^* c_k$, the evolution of these centers is given by

$$c_{k+1}^a = (A - G^*C - BL^*)c_k^a + G^* y_k^r \tag{27a}$$

$$c_{k+1}^r = (A - G^*C - BL^*)c_k^r + G^* y_k^r \tag{27b}$$

starting at $c_{k_1}^a = c_{k_1}$ and $c_{k_1}^r = c_{k_0}$.

Consequently, for $k \in K_{REP}$, the attacked residuals are governed by the equation

$$r_k^a = r_k^r + C(A - G^*C - BL^*)^{k-k_1}(c_{k_1}^r - c_{k_1}^a). \tag{28}$$

In particular, we focus on the case in which the attack is undetectable for an unprotected system like the one represented in Fig. 1. Then, the following assumption will be followed.

Assumption 3: The attack is undetectable by the passive detector (25). That is, for all $k \in K_{REP}$, the residual signal satisfies

$$r_k^r + C(A - G^*C - BL^*)^{k-k_1}(c_{k_1}^r - c_{k_1}^a) \in \mathcal{R}^H.$$

Note that the consideration of Assumption 3 motivates the injection of an additive watermark signal ξ_k in the system inputs to detect the attack. In this regard, conversely to the stochastic approaches that inject an exogenous Gaussian signal [11], in what follows signal ξ_k will be assumed to be bounded within the known zonotope $\xi_k = \langle 0, H_\xi \rangle$. Accordingly, the first step is to analyze the impact on the system performance that entails the injection of the suboptimal signal $u_k = u_k^* + \xi_k$.

C. Performance Loss Induced By A Zonotopically-Bounded Watermark Signal

For the case where the optimal control law derived in Sec-

tion IV is extended with the inclusion of the additive signal $\xi_k \in \langle 0, H_\xi \rangle$, the following theorem can be established.

Theorem 3: The injection of the control law $u_k = -L_k^* c_k + \xi_k$, with L_k^* in (15) and $\xi_k \in \langle 0, H_\xi \rangle$, yields the optimal cost-to-go

$$\bar{V}_k(x_k) = Q[c_k^T S_k c_k] + \bar{s}_k$$

where $\bar{s}_k$ is described by the backward recursion

$$\bar{s}_k = \bar{s}_{k+1} + Tr[W\check{P}_k + S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})] + Tr[(U + B^T S_{k+1} B)P_\xi]$$

starting at $\bar{s}_N = Tr[W_f \check{P}_N]$, with S_k defined in (14a) and with $P_\xi = H_\xi H_\xi^T$.

Proof: The addition of the exogenous signal $\xi_k \in \langle 0, H_\xi \rangle$ implies that input signal satisfies $u_k = u_k^* + \xi_k \in \langle u_k^*, H_\xi \rangle$. Therefore, the proof is obtained by following the same steps as for Theorem 1 but taking into consideration that for this case:

$$Q[u_k^T U u_k] = Q[(u_k^* + \xi_k)^T U (u_k^* + \xi_k)] = u_k^{*T} U u_k^* + Tr[U P_\xi]$$

with $P_\xi = H_\xi H_\xi^T$, and that the center of the state-estimator (6a) now satisfies

$$c_{k+1} \in \langle (Ac_k + Bu_k^*), [G_k^* C H_k,\ G_k^* F,\ B H_\xi] \rangle. \quad \blacksquare$$

According to Theorem 3, since the solution of the infinite horizon control problem is given by the limit of the finite horizon solution, the new performance loss obtained with the inclusion of the bounded watermarking signal J_∞^{wm} can be assessed as

$$\begin{aligned} J_\infty^{wm} &= Tr[WP_\infty + S_\infty(G^*(CP_\infty C^T + Q_v)G^{*T}] \\ &\quad + Tr[(U + B^T S_\infty B)P_\xi] \\ &= J_\infty + Tr[(U + B^T S_\infty B)P_\xi] \\ &= J_\infty + \Delta J. \end{aligned} \tag{29}$$

It must be highlighted that the extra term in the cost function

$$\Delta J = Tr[(U + B^T S_\infty B)P_\xi] = \|H_\xi\|^2_{F, W_\xi} \tag{30}$$

with $W_\xi = U + B^T S_\infty B$, matches the performance loss that induces the injection of a random Gaussian variable $\xi_k \sim \mathcal{N}(0, P_\xi)$ in an LQG control scheme [11]. That is, there is a parallelism in the effect that causes in the system performance to address the watermarking design by means of injecting an exogenous Gaussian signal or through its zonotopic counterpart.

Remark 5: Notice that the analogy discussed above is based on the parallelism between a zonotopically-bounded random variable $x \in \langle 0, H \rangle$ (i.e., for an unspecified probability density function the support $\mathcal{S}_x$ of x satisfies $\mathcal{S}_x \subseteq \langle 0, H \rangle$ (cf. [21])) and a Gaussian random variable $x' \sim \mathcal{N}(0, Q)$ whose covariance matrix satisfies $Q = HH^T$. Nevertheless, an explicit relationship between the bounding zonotope $\langle 0, H \rangle$ of x and the confidence ellipsoids of x' (which has an unbounded support) is not possible due to the fact that multiple zonotopes share the same covariation matrix. Therefore, the selection of one uncertainty characterization over the other should be based on its capability to reliably describe the uncertain data at the modeling stage.

VI. Zonotopically Bounded Watermark Signal Design

This section deals with the design of a watermark signal that can adopt any value within a given zonotope. Working under a set-based paradigm, this watermark signal must be designed in such a way that it guarantees the attack detection (i.e., such that enforces $r_k^a \notin \mathcal{R}_H$). To that end, the proposed scheme takes advantage of the fact that the observability of a signal that is known by the defender is lost during the replay phase of the attack. Hence, this fact is used in order to destabilize the estimation error of such known signal, for which, a zonotopic observer provides explicit bounds.

In order to implement the scheme described in the preceding paragraph, the closed-loop dynamics of the system are gathered in the model

$$\begin{aligned} \bar{x}_{k+1} &= \bar{A}\bar{x}_k + \bar{B}\xi_k + \bar{E}\bar{w}_k \\ y_k &= \bar{C}\bar{x}_k + Fv_k \end{aligned} \tag{31}$$

with $\bar{x}_k = [x_k^T,\ e_k^T]^T$, $\bar{w}_k = [w_k^T,\ v_k^T]^T$, the new system matrices

$$\bar{A} = \begin{bmatrix} A - BL^* & BL^* \\ 0 & A - G^*C \end{bmatrix}, \quad \bar{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}$$
$$\bar{E} = \begin{bmatrix} E & 0 \\ E & -G^*F \end{bmatrix}, \quad \bar{C} = \begin{bmatrix} C & 0 \end{bmatrix}$$

and the initial conditions $\bar{x}_0 = [x_0^T,\ e_0^T]^T \in \langle \bar{c}_0, \bar{H}_0 \rangle$ where

$$\bar{c}_0 = \begin{bmatrix} c_0 \\ 0 \end{bmatrix}, \quad \bar{H}_0 = \begin{bmatrix} H_0 & 0 \\ 0 & H_0 \end{bmatrix}.$$

Henceforth, the watermark signal $\xi_k \in \mathbb{R}^{n_u}$ is defined as the difference

$$\xi_k = \psi - c_k^\psi \tag{32}$$

where, $\psi \in \mathbb{R}^{n_u}$ is a nonzero constant vector set by the defender (i.e., $\psi \neq 0$), and $c_k^\psi \in \mathbb{R}^{n_u}$ is the center of a zonotopic observer that will be introduced later.

Let the constant vector ψ be rewritten as

$$\psi = M\psi - (M - I_{n_u})\psi \tag{33}$$

for any given matrix $M \in \mathbb{R}^{n_u \times n_u}$ which has been included as a tuning parameter. Thus, the injection of the exogenous signal (32) in the closed-loop system (31), can be written as

$$\begin{bmatrix} \bar{x}_{k+1} \\ \psi \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{B} \\ 0 & M \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \psi \end{bmatrix} - \begin{bmatrix} \bar{B}c_k^\psi \\ (M - I_{n_u})\psi \end{bmatrix} + \begin{bmatrix} \bar{E} \\ 0 \end{bmatrix} \bar{w}_k \tag{34a}$$

$$y_k = \begin{bmatrix} \bar{C} & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \psi \end{bmatrix} + Fv_k \tag{34b}$$

$$\begin{bmatrix} \bar{x}_0 \\ \psi \end{bmatrix} \in \left\langle \begin{bmatrix} \bar{c}_0 \\ c_0^\psi \end{bmatrix}, \begin{bmatrix} \bar{H}_0 \\ 0 \end{bmatrix} \right\rangle, \quad c_0^\psi = \psi. \tag{34c}$$

A. Design of an Extended State Estimator

In order to design a new observer for estimating the value of ψ, the time-varying matrix $\bar{G}_k \in \mathbb{R}^{(2n_x + n_u) \times n_y}$ is introduced in

(34a) yielding the equivalent system

$$\begin{bmatrix} \bar{x}_{k+1} \\ \psi \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{B} \\ 0 & M \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \psi \end{bmatrix} - \begin{bmatrix} \bar{B} c_k^{\psi} \\ (M - I_{n_u})\psi \end{bmatrix} + \begin{bmatrix} \bar{E} \\ 0 \end{bmatrix} \bar{w}_k + \bar{G}_k (y_k - \begin{bmatrix} \bar{C} & 0 \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \psi \end{bmatrix} - F v_k). \tag{35}$$

At this point, the following notation is introduced:

$$\tilde{x}_k = \begin{bmatrix} \bar{x}_k \\ \psi \end{bmatrix}, \quad \tilde{c}_k = \begin{bmatrix} \bar{c}_k \\ c_k^{\psi} \end{bmatrix}, \quad \tilde{H}_0 = \begin{bmatrix} \bar{H}_0 \\ 0 \end{bmatrix} \tag{36}$$

in such a way that, at the initial time instant, $\tilde{x}_0 \in \langle \tilde{c}_0, \tilde{H}_0 \rangle$.

Theorem 4: Given the system (34a) and (34b) with the initial conditions satisfying the inclusion (34c). By recursively defining the center and the generators matrix

$$\tilde{c}_{k+1} = (\hat{A} - \bar{G}_k \tilde{C}) \tilde{c}_k - \tilde{M} \psi + \bar{G}_k y_k \tag{37a}$$

$$\tilde{H}_{k+1} = \left[(\tilde{A} - \bar{G}_k \tilde{C}) \bar{H}_k, \tilde{E}, -\bar{G}_k F \right] \tag{37b}$$

with matrices

$$\hat{A} = \begin{bmatrix} \bar{A} & 0 \\ 0 & M \end{bmatrix}, \ \tilde{A} = \begin{bmatrix} \bar{A} & \bar{B} \\ 0 & M \end{bmatrix}, \ \tilde{C} = \begin{bmatrix} \bar{C} & 0 \end{bmatrix}$$

$$\tilde{E} = \begin{bmatrix} \bar{E} \\ 0 \end{bmatrix}, \qquad \tilde{M} = \begin{bmatrix} 0 \\ M - I_{n_u} \end{bmatrix} \tag{38}$$

then the inclusion property $\tilde{x}_k \in \langle \tilde{c}_k, \tilde{H}_k \rangle$ holds for all $k \geq 0$.

Proof: Assuming that the inclusion is satisfied at time k, which according to (34c) is true for $k = 0$, then using (35) and the notation introduced in (36), it follows that:

$$\tilde{x}_k \in \langle \tilde{c}_k, \tilde{H}_k \rangle \implies \tilde{x}_{k+1} \in \langle \tilde{c}_{k+1}, \tilde{H}_{k+1} \rangle$$

where

$$\begin{aligned} \langle \tilde{c}_{k+1}, \tilde{H}_{k+1} \rangle = & \begin{bmatrix} \bar{A} & \bar{B} \\ 0 & M \end{bmatrix} \langle \tilde{c}_k, \tilde{H}_k \rangle \oplus \begin{bmatrix} 0 & -\bar{B} \\ 0 & 0 \end{bmatrix} \langle \tilde{c}_k, 0 \rangle \\ & \oplus \begin{bmatrix} 0 \\ M - I_{n_u} \end{bmatrix} \langle \psi, 0 \rangle \oplus \begin{bmatrix} \bar{E} \\ 0 \end{bmatrix} \langle 0, I_{n_w + n_v} \rangle \\ & \oplus \bar{G}_k \langle y_k, 0 \rangle \oplus - \bar{G}_k \begin{bmatrix} \bar{C} & 0 \end{bmatrix} \langle \tilde{c}_k, \tilde{H}_k \rangle \\ & \oplus - \bar{G}_k F \langle 0, I_{n_v} \rangle. \end{aligned}$$

Therefore, using (3) and defining the set of matrices in (38), the expressions for the center (37a) and generators matrix (37b) are retrieved, and thus the inclusion property is satisfied at $k+1$. This gives the proof by induction. ■

Accordingly, from Theorem 4 it follows that:

$$\tilde{x}_k \in \langle \tilde{c}_k, \tilde{H}_k \rangle \implies \tilde{x}_k - \tilde{c}_k \in \langle 0, \tilde{H}_k \rangle$$

and, as a result, if the projection matrix $N = [0 \ I_{n_u}] \in \mathbb{R}^{n_u \times (2n_x + n_u)}$ is defined, then the watermarking signal (32) satisfies

$$\xi_k = N(\tilde{x}_k - \tilde{c}_k) \in N \langle 0, \tilde{H}_k \rangle = \langle 0, N \tilde{H}_k \rangle. \tag{39}$$

Additionally, with regard to the design of a matrix $\bar{G}_k$ that minimizes the performance loss induced by the signal (39) (cf. Section V-C), the following theorem is introduced.

Theorem 5: Given an SPD matrix W_ξ and the generators matrix $\tilde{H}_{k+1}$ as in (37b). Then, the additional term $\Delta J = \|N \tilde{H}_{k+1}\|^2_{F, W_\xi}$ induced by the injection of signal ξ_k is minimized for the observer matrix $\bar{G}_k^* = \tilde{A} \tilde{P}_k \tilde{C}^T (\tilde{C} \tilde{P}_k \tilde{C}^T + Q_v)^{-1}$, with N being the projection matrix in (39) and $\tilde{P}_k = \tilde{H}_k \tilde{H}_k^T$.

Proof: From Definition 4 it follows that:

$$\|\langle N \tilde{H}_{k+1} \rangle\|^2_{F, W_\xi} = Tr[\tilde{H}_{k+1}^T \tilde{W} \tilde{H}_{k+1}]$$

with matrix $\tilde{W}$ defined as

$$\tilde{W} = N^T W_\xi N = \begin{bmatrix} 0 & 0 \\ 0 & W_\xi \end{bmatrix}.$$

Hence, since $Tr[\tilde{H}_{k+1}^T \tilde{W} \tilde{H}_{k+1}]$ is convex with respect to $\bar{G}_k$, following the same steps as in the proof of Theorem 5 in [27], it follows that the matrices $\tilde{G}_k^* = \arg\min_{\tilde{G}} Tr[\tilde{H}_{k+1}^T \tilde{W} \tilde{H}_{k+1}]$ must satisfy the equation:

$$\tilde{W} \bar{G}_k^* = \tilde{W} \tilde{A} \tilde{P}_k \tilde{C}^T (\tilde{C} \tilde{P}_k \tilde{C}^T + Q_v)^{-1}. \tag{40}$$

Then, by taking into consideration that $\tilde{W}$ is symmetric positive semi-definite, $\bar{G}_k^* = \tilde{A} \tilde{P}_k \tilde{C}^T (\tilde{C} \tilde{P}_k \tilde{C}^T + Q_v)^{-1}$ is one solution that satisfies (40). ■

B. Watermarked System Stability

Analogously to Section IV-B, in the sequel the fixed steady-state observer gain $\bar{G}^*$ is considered. Next, the stability of the closed-loop system (31) subject to the injection of the watermark signal ξ_k is analyzed. To that end, through the comparison of (34a) with (37a), it can be seen that the estimation error produced by the new observer is governed by the dynamics

$$\tilde{e}_{k+1} = (\tilde{A} - \bar{G}^* \tilde{C}) \tilde{e}_k + \tilde{E} \bar{w}_k - \bar{G}^* F v_k \tag{41}$$

with $\tilde{e}_k = [(\bar{x}_k - \bar{c}_k)^T \ (\psi - c_k^{\psi})^T]^T$.

Consequently, if the evolution of the closed-loop system (31) and the new estimation error (41) are combined in a single dynamical system, this has the form

$$\begin{bmatrix} \bar{x}_{k+1} \\ \tilde{e}_{k+1} \end{bmatrix} = \begin{bmatrix} \bar{A} & \bar{B} N \\ 0 & \tilde{A} - \bar{G}^* \tilde{C} \end{bmatrix} \begin{bmatrix} \bar{x}_k \\ \tilde{e}_k \end{bmatrix} + \begin{bmatrix} \bar{E} \\ \tilde{E} \end{bmatrix} \bar{w}_k + \begin{bmatrix} 0 \\ -\bar{G}^* F \end{bmatrix} v_k \tag{42}$$

with an upper triangular state transition matrix. Therefore, the stability of the watermarked system (42) is related with the eigenvalues of $\bar{A}$ (which is associated to an asymptotically stable system), and the eigenvalues of $\tilde{A} - \bar{G}^* \tilde{C}$.

Note that the design of an observer gain $\bar{G}^*$ such that $\tilde{A} - \bar{G}^* \tilde{C}$ is asymptotically stable depends on the detectability of the pair $(\tilde{A}, \tilde{C})$. In this regard, the following mild sufficient condition can be stated.

Proposition 1: A sufficient condition for the detectability of the pair $(\tilde{A}, \tilde{C})$, is that there is a column of matrix $\bar{B}$, denoted as $\bar{B}_{(:,i)}$, that satisfies

$$rank(\bar{C}[\bar{B}_{(:,i)}, \bar{A} \bar{B}_{(:,i)}, \ldots, \bar{A}^{2n_x - 2} \bar{B}_{(:,i)}]) \geq 1$$

where $2n_x$ is the dimension of the square matrix $\bar{A}$.

Proof: Given a matrix $\tilde{A}$ defined in (38), then from the asymptotic stability of $\bar{A}$ it follows that it is only required to guarantee the observability of the modes associated with matrix M. To that end, consider that the signal ξ_k is introduced only through the i-th column of matrix $\bar{B}$ denoted as

$\bar{B}_{(:,i)}$. Then, the observability matrix $O(\tilde{A},\tilde{C})$ has the structure

$$O(\tilde{A},\tilde{C}) = \begin{bmatrix} O(\bar{A},\bar{C}) & O_{12} \\ \bar{C}\bar{A}^{n_x} & O_{22} \end{bmatrix} \in \mathbb{R}^{n_y(2n_x+1)\times(2n_x+1)}$$

with

$$O_{22} = \bar{C}\bar{A}^{2n_x-1}\bar{B}_{(:,i)} + \bar{C}\sum_{j=1}^{2n_x-1}\bar{A}^{2n_x-1-j}\bar{B}_{(:,i)}M^j \in \mathbb{R}^{n_y}.$$

Hence, if $rank(\bar{C}[\bar{B}_{(:,i)}, \bar{A}\bar{B}_{(:,i)}, \ldots, \bar{A}^{2n_x-2}\bar{B}_{(:,i)}]) \geq 1$, then the term O_{22} can be modified by means of the parameters M, thus allowing to impose the independence of the last column of O_2 from the first $2n_x$ columns. Therefore, the mode associated with matrix M is observable, and thus $(\tilde{A},\tilde{C})$ detectable. ■

Note that, from [43, Theorem 10.3], with Q_v SPD and matrix $\tilde{A}$ whose eigenvalues depend on the asymptotically stable matrix $\bar{A}$ (eigenvalues strictly inside the unit disk) and the arbitrary matrix M, then Proposition 1 provides a sufficient condition for the applicability of the proposed watermarking method.

C. Attack Detection

Here the detectability of the replay attack under the proposed watermark signal is analyzed. To that end, let Ψ denote the mRPI set for the estimation error generated by the extended observer (41), and let $\langle 0, \tilde{H}^{ss}\rangle \supseteq \Psi$ denote an RPI zonotopic outer-approximation of Ψ. Furthermore, note that, if the matrix $\bar{G}^*$ is designed according to Theorem 5, then from Theorem 2 it follows that $\|\Psi\|_F^2 = Tr[P_\infty^\Psi]$, with P_∞^Ψ being the solution of the corresponding DARE.

Besides, taking advantage of the fact that the constant vector ψ is known by the defender, the watermark signal ξ_k can also be used in order to assess the system operation. Hence, the following can be stated for the steady-state:

$$\begin{cases} \xi_k \in \langle 0, N\tilde{H}^{ss}\rangle & \Longrightarrow \text{ Healthy operation,} \\ \text{otherwise} & \Longrightarrow \text{ Something is wrong.} \end{cases} \quad (43)$$

The blue dashed box in Fig. 1 shows the structure of watermarking scheme and its interaction with the original control loop. Furthermore, following the notation presented in Section V, during the replay phase of the attack the watermark signal ξ_k^a can be written as

$$\xi_k^a = \psi - c_k^{\psi,a} = \psi - c_k^{\psi,r} + (c_k^{\psi,r} - c_k^{\psi,a}) = \xi_k^r + (c_k^{\psi,r} - c_k^{\psi,a}) \quad (44)$$

where $c_k^{\psi,r}$ and $c_k^{\psi,a}$ are the center of the estimation of ψ during the record and replay phases, respectively. Whereas, extending steady-state condition introduced in Assumption 2 to the extended estimation error (41), the signal ξ_k^r satisfies

$$\xi_k^r = \psi - c_k^{\psi,r} \in \langle 0, N\tilde{H}^{ss}\rangle, \quad \forall k \in K_{REP}. \quad (45)$$

The following proposition shows that the watermark signal ξ_k can be tuned to detect replay attacks.

Proposition 2: If any of the eigenvalues of matrix M in (33) is strictly outside the unit disk, then the watermark signal during the replay phase ξ_k^a will ultimately exit the set $\langle 0, N\tilde{H}^{ss}\rangle$.

Proof: From (44), and taking into consideration (45), the signal ξ_k^a depends on the difference $c_k^{\psi,r} - c_k^{\psi,a}$, which can be rewritten as $c_k^{\psi,r} - c_k^{\psi,a} = N(\tilde{c}_k^r - \tilde{c}_k^a)$.

On the other hand, the center of the estimator (37a) during the record and replay phases evolves as

$$\tilde{c}_{k+1}^r = (\hat{A} - \bar{G}^*\tilde{C})\tilde{c}_k^r - \tilde{M}\psi + \bar{G}^* y_k^r \quad (46a)$$

$$\tilde{c}_{k+1}^a = (\hat{A} - \bar{G}^*\tilde{C})\tilde{c}_k^a - \tilde{M}\psi + \bar{G}^* y_k^r \quad (46b)$$

starting at $\tilde{c}_{k_1}^r = [\bar{c}_{k_0}^T \; c_{k_0}^{\psi,T}]^T$ and $\tilde{c}_{k_1}^a = [\bar{c}_{k_1}^T \; c_{k_1}^{\psi,T}]^T$.

Recalling matrix $\hat{A}$ in (38) and splitting the matrix $\bar{G}^*$ into the blocks

$$\hat{A} = \begin{bmatrix} \bar{A} & 0 \\ 0 & M \end{bmatrix}, \quad \bar{G}^* = \begin{bmatrix} \bar{G}_1 \\ \bar{G}_2 \end{bmatrix}$$

then from the comparison of (46a) and (46b) it follows that the center difference between phases evolves according to:

$$\begin{bmatrix} \bar{c}_{k+1}^r - \bar{c}_{k+1}^a \\ c_{k+1}^{\psi,r} - c_{k+1}^{\psi,a} \end{bmatrix} = \begin{bmatrix} \bar{A} - \bar{G}_1\bar{C} & 0 \\ -\bar{G}_2\bar{C} & M \end{bmatrix} \begin{bmatrix} \bar{c}_k^r - \bar{c}_k^a \\ c_k^{\psi,r} - c_k^{\psi,a} \end{bmatrix}. \quad (47)$$

Therefore, since the state matrix of (47) is lower-triangular, if matrix M has any eigenvalue outside the unit disk, then the dynamics of $c_k^{\psi,r} - c_k^{\psi,a}$ are unstable, thus forcing that ξ_k^a will ultimately come out of the healthy set $\langle 0, N\tilde{H}^{ss}\rangle$. ■

Remark 6: In the moment that the attack is detected, the signal ξ_k^a should stop being injected so as not to affect the system with its exponential growth.

VII. Numerical Example

The quadruple-tank system [47] is a well-known benchmark used to evaluate control and supervision strategies, and it is also employed to evaluate cyber-attack detection techniques in [5], [7]. The nonlinear dynamics of the system are described as follows:

$$\begin{cases} \dfrac{dh_1}{dt} = -\dfrac{a_1}{A_1}\sqrt{2gh_1} + \dfrac{a_3}{A_1}\sqrt{2gh_3} + \dfrac{\gamma_1 k_1}{A_1}v_1 + e_1 w_1 \\ \dfrac{dh_2}{dt} = -\dfrac{a_2}{A_2}\sqrt{2gh_2} + \dfrac{a_4}{A_2}\sqrt{2gh_4} + \dfrac{\gamma_2 k_2}{A_2}v_2 + e_1 w_1 + e_2 w_2 \\ \dfrac{dh_3}{dt} = -\dfrac{a_3}{A_3}\sqrt{2gh_3} + \dfrac{(1-\gamma_2)k_2}{A_3}v_2 + e_1 w_1 + e_3 w_3 \\ \dfrac{dh_4}{dt} = -\dfrac{a_4}{A_4}\sqrt{2gh_4} + \dfrac{(1-\gamma_1)k_1}{A_4}v_1 + e_1 w_1 + e_4 w_4 \\ y_1 = k_c h_1 + f_1 v_1, \;\; y_2 = k_c h_2 + f_2 v_2 \end{cases} \quad (48)$$

where v_j, for $j \in 1,2$, is the control input such that $k_j v_j$ is the j-th input flow rate, $h_i \in [0,20]$ cm denotes the water level of the i-th tank, A_i denotes the cross-section area of the i-th tank, a_i denotes the cross-section area of the i-th outlet hole, g is the gravity acceleration, and $\gamma_j \in [0,1]$ is the position of the j-th valve. The values of these parameters are given in Table I. The model (48) is linearized around the operating point indicated by means of the superscript ^{o} using Euler discretization with a sampling time $T_s = 1$ s. A linear discrete-time model as (4) can be obtained, describing the dynamics of the variations $\Delta h_{k,i} = h_{k,i} - h_{k,i}^o$ and $\Delta v_{k,i} = v_{k,i} - v_{k,i}^o$, such that

TABLE I
QUADRUPLE TANK PROCESS PARAMETERS

Parameter	Value	Unit
$A_1 = A_3$	28	cm^2
$A_2 = A_4$	32	cm^2
$a_1 = a_3$	0.071	cm^2
$a_2 = a_4$	0.05	cm^2
k_c	0.5	V/cm
g	981	cm/s^2
(h_1^o, h_2^o)	(12.4, 12.7)	cm
(h_3^o, h_4^o)	(1.8, 1.4)	cm
(v_1^o, v_2^o)	(3, 3)	V
(k_1^o, k_2^o)	(3.33, 3.35)	$cm^3/V\ s$
(γ_1^o, γ_2^o)	(0.7, 0.6)	–
e_1	0.05	–
e_2	0.01	–
$e_3 = e_4$	0.02	–
f_1	0.03	–
f_2	0.01	–

$$x_k = \begin{bmatrix} \Delta h_{k,1} \\ \Delta h_{k,2} \\ \Delta h_{k,3} \\ \Delta h_{k,4} \end{bmatrix}, \quad u_k = \begin{bmatrix} \Delta v_{k,1} \\ \Delta v_{k,2} \end{bmatrix}, \quad y_k = \begin{bmatrix} \Delta y_{k,1} \\ \Delta y_{k,2} \end{bmatrix}$$

$$A = \begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{A_3}{A_1 T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{A_4}{A_2 T_2} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix}, \quad T_i = \frac{A_i}{a_i}\sqrt{\frac{2h_i^o}{g}}$$

$$B = \begin{bmatrix} \frac{\gamma_1^o k_1^o}{A_1} & 0 \\ 0 & \frac{\gamma_2^o k_2^o}{A_2} \\ 0 & \frac{(1-\gamma_2^o)k_2^o}{A_3} \\ \frac{(1-\gamma_1^o)k_1^o}{A_4} & 0 \end{bmatrix}, \quad E = \begin{bmatrix} e_1 & 0 & 0 & 0 \\ e_1 & e_2 & 0 & 0 \\ e_1 & 0 & e_3 & 0 \\ e_1 & 0 & 0 & e_4 \end{bmatrix}$$

$$C = [k_c I_2 \ \ 0], \quad F = \begin{bmatrix} f_1 & 0 \\ 0 & f_2 \end{bmatrix}.$$

In addition, it has been considered that system disturbances w_k and measurement noise v_k are uniformly distributed random variables bounded within the zonotopes $w_k \in \langle 0, I_4 \rangle$ and $v_k \in \langle 0, I_2 \rangle$, for all $k \in \mathbb{N}$. The process is controlled by means of an optimal LQR controller designed for the weighting matrices $W = 10 \cdot I_4$ and $Q = I_2$. The fixed steady-state gain of the controller is

$$L^* = \begin{bmatrix} 2.2171 & 0.0004 & 0.1650 & 0.0744 \\ -0.0016 & 2.4915 & 1.0060 & 0.4850 \end{bmatrix}.$$

Besides, by taking into consideration the covariation matrices $Q_w = EE^T$ and $Q_v = FF^T$, the obtained steady-state gain of the ZKF is

$$G^* = \begin{bmatrix} 0.5106 & 0.1399 & 0.5472 & 0.1060 \\ 1.2686 & 1.6852 & 1.1097 & 1.6209 \end{bmatrix}^T.$$

At this point, it must be highlighted that the eigenvalues of matrix $A - G^*C - BL^*$ are $\{-0.1739, 0.4005, 0.9513 \pm 0.0278i\}$, that is, it is asymptotically stable, and thus there is no guarantee that the replay attack can be detected (cf. (28)). In other words, in the remainder of this section, Assumption 3 concerning the attack undetectability prevails thus motivating the injection of a watermark signal in the control loop.

A. Performance Assessment

Concerning the estimation error $e_k = x_k - c_k$, from Theo rem 2 it follows that the size of its mRPI set Φ can be computed directly by solving the corresponding DARE. By doing so, the obtained F-radius is $\|\Phi\|_F = \sqrt{Tr[P_\infty]} = 0.1393$, with P_∞ being the unique positive definite solution of the DARE. Consequently, if the mRPI set Φ could be used in order to bound the estimation error in the steady-state, applying (22) it follows that the optimal infinite horizon cost function that could be obtained would be $J_\infty = 0.9275$.

Due to the difficulties of computing an exact representation of the Φ, different RPI outer-approximations of such mRPI set will be computed below. In this regard, an initial low order RPI zonotopic set $\mathcal{Z}_0^{ss} = \langle 0, H_0^{ss} \rangle$ is computed as detailed in Appendix C. The obtained set is a parallelotope, that is, a first order zonotope, with the following generators matrix:

$$H_0^{ss} = \begin{bmatrix} 0.0894 & 0.0256 & -0.3773 & 0.3080 \\ 0.0928 & -0.0238 & 0.0413 & -0.0335 \\ 0.0898 & -0.2360 & -0.9798 & -0.0529 \\ 0.0935 & -0.5073 & 0.1513 & -0.0000 \end{bmatrix}.$$

The initial RPI set $\mathcal{Z}_0^{ss}$ is now forward propagated obtaining successively tighter RPI over-approximations of the mRPI set Φ. In this regard, Table II(a) shows the zonotope order and the F-radius of the over-approximations obtained for different iteration values. Besides, the value of the cost function for the infinite horizon problem is also shown.

After analyzing the data shown in Table II(a), it can be seen how the F-radius of the successive over-approximations converges to the previously computed $\|\Phi\|_F = 0.1393$ value (shown in red). However, this is achieved at the cost of increasing the complexity of the zonotope as reflected in the second column. Note that, the bad performance index obtained with the initial set motivates its further propagation in search of sets whose F-radius, and hence the imposed performance loss is closer to the optimal value.

Furthermore, it must be highlighted that the computation of the sets shown in Table II(a) is done offline. Then, the only online computation is the evaluation of the condition $r_k \in \mathcal{R}_i^H$, with $\mathcal{R}_i^H = C\mathcal{Z}_i^{ss} \oplus F\mathcal{V} = \langle 0, [CH_i^{ss}, F] \rangle$. Testing whether or not a point belongs to a zonotopic set can be efficiently done by solving a constraint satisfaction problem.

B. Watermark Signal Design

In order to design a watermark signal like the one presented in Section VI, the tuning matrix M must be defined. In this regard, the following matrix M is introduced:

$$M = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.05 \end{bmatrix}$$

TABLE II
ZONOTOPIC RPI APPROXIMATIONS OF THE MRPI SET

(a) State estimation

Iteration	Order	F-radius	$\tilde{J}_\infty$
H_1^{ss}	1	1.2544	19.5451
H_5^{ss}	7	0.8034	7.7927
H_{10}^{ss}	14.5	0.4994	3.3956
H_{20}^{ss}	29.5	0.2368	1.3144
H_{30}^{ss}	44.5	0.1610	0.9958
H_{50}^{ss}	74.5	0.1405	0.9307
H_{75}^{ss}	112	0.1394	0.9276
H_{100}^{ss}	149.5	0.1393	0.9275
Φ_∞	∞	0.1393	0.9275

(b) Watermark signal

Iteration	Order	F-radius	$\Delta\tilde{J}$
$\tilde{H}_1^{ss}$	1	10.9688	127.7466
$\tilde{H}_5^{ss}$	4.2	9.0130	86.0918
$\tilde{H}_{10}^{ss}$	8.2	7.1512	54.0752
$\tilde{H}_{20}^{ss}$	16.2	4.5690	22.0604
$\tilde{H}_{30}^{ss}$	24.2	2.9530	9.2396
$\tilde{H}_{50}^{ss}$	40.2	1.2805	1.7299
$\tilde{H}_{75}^{ss}$	60.2	0.6131	0.3662
$\tilde{H}_{100}^{ss}$	80.2	0.5059	0.2361
Ψ_∞	∞	0.4929	0.2223

so as to meet Proposition 2. Then, following the development presented in Section VI, the optimal steady-state value of the gain $\bar{G}^*$ of the new observer is:

$$\bar{G}^* = \begin{bmatrix} \bar{G}_a & \bar{G}_b \end{bmatrix}^T$$

$$\bar{G}_a = \begin{bmatrix} 0.6203 & 0.1388 & 0.1039 & 0.0202 & 0.1204 \\ 1.2684 & 1.8383 & 1.8712 & -0.238 & 0.0427 \end{bmatrix}$$

$$\bar{G}_b = \begin{bmatrix} 0.0037 & 0.1951 & -0.014 & 0.8998 & -0.130 \\ 0.1566 & -0.030 & 0.1739 & 0.2774 & 4.1170 \end{bmatrix}.$$

Concerning the new estimation error, the F-radius of its mRPI set Ψ is $\|\Psi\|_F = \sqrt{Tr[P_\infty^\Psi]} = 0.4929$, with P_∞^Ψ being the unique positive definite solution of the corresponding DARE. Hence, by defining the projection matrix $N = [0_{2\times 8},\ I_2]$ and following the developments presented in Section V-C, the performance loss induced by a watermark signal confined within a zonotope with covariance $P_\xi = NP_\infty^\Psi N^T$ is given by

$$\Delta J = Tr[(U + B^T S_\infty B)P_\xi] = 0.2223.$$

Next, similarly to Section VII-A several RPI outer-approximations to Ψ are computed starting from an initial first order zonotope $\tilde{\mathcal{Z}}_0^{ss} = \langle 0, \tilde{H}_0^{ss} \rangle$. In this regard, Table II(b) presents the zonotope order and the F-radius of the over-approximations obtained for different iteration values. Note that given the set RPI set $\tilde{\mathcal{Z}}_i^{ss} = \langle 0, \tilde{H}_i^{ss} \rangle$, the healthy watermark signal ξ_k is guaranteed to converge within $\langle 0, N\tilde{H}_i^{ss} \rangle$ and, once it is inside, it will only exit that set in case the system in under attack. Consequently, the fourth column of Table II(b) shows for each iteration the performance loss ΔJ induced by a watermark signal constrained to $\xi_k \in \langle 0, N\tilde{H}_i^{ss} \rangle$ and thus with the covariation matrix $P_\xi = N\tilde{H}_i^{ss}\tilde{H}_i^{ss,T}N^T$.

C. Attack Simulation

A replay attack has been simulated with the time windows

$$\mathcal{K}_{REC} = [300,\ 500], \quad \mathcal{K}_{REC} = [700,\ 900].$$

Furthermore, the selected RPI over-approximation for the ZKF is $\mathcal{Z}_{20}^{ss}$ while for the state estimator used in the watermarking generation the selected RPI over-approximation is $\tilde{\mathcal{Z}}_{50}^{ss}$. The temporal evolution of the watermark signal during the attack scenario is represented in Fig. 2. In this figure, it can be seen how after the start of the replay phase at $k = 700$, the watermark signal starts growing exponentially until the time instant $k = 737$, when the condition $\xi_k^a \in N\tilde{\mathcal{Z}}_{50}^{ss}$ is no longer satisfied and thus the attack is detected. The watermark signal stops being injected after the attack is detected.

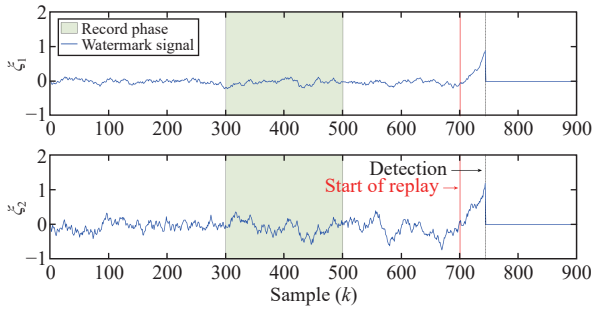


Fig. 2. Temporal evolution of the watermark signal.

On the other hand, Fig. 3 shows the effect caused on the system outputs by the injection of the watermark signal. It can be seen how the performance of the system is hardly degraded before the replay phase and how after the start of this phase the effect of the increasing watermark signal becomes more evident. It should be highlighted that since the control loop is affected by the data replay, the stability of the plant after the attack detection, at $k = 737$, is due solely to the fact that the open-loop system is stable.

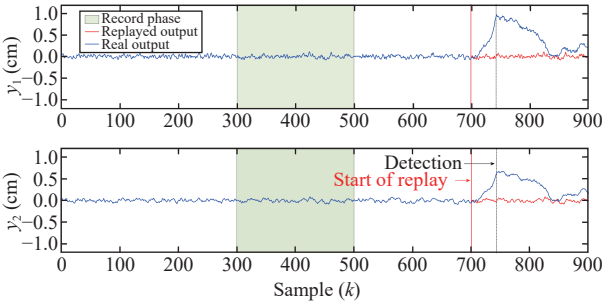


Fig. 3. Temporal evolution of the system outputs.

The mean times (t_m) required in the simulations to monitor the operation of the system online are[1]

[1] Laptop (Intel i7 1.8 GHz, 16 GB RAM) running Windows 10; optimization using Cplex 12.8 [48].

1) Evaluation of $\xi_k \in N\tilde{\mathcal{Z}}_{50}^{ss} = \langle 0, N\tilde{H}_{50}^{ss}\rangle$: $t_m = 8.9$ ms.

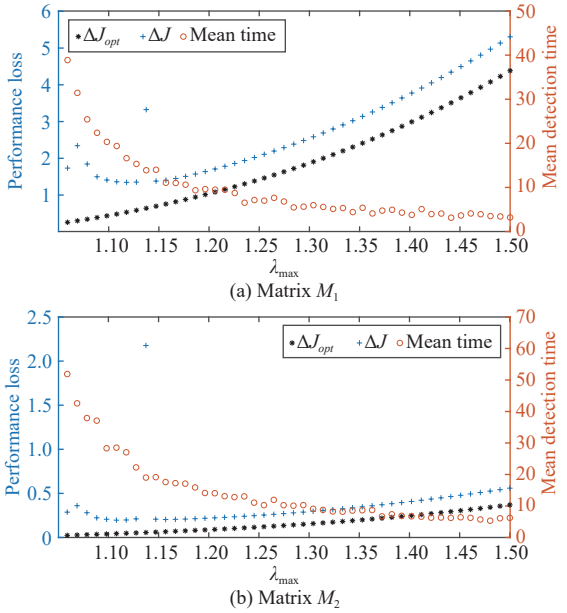
2) Evaluation of $r_k \in \mathcal{R}^H = \langle 0, [CH_{20}^{ss},\ F]\rangle$: $t_m = 6.3$ ms.

and thus well below the $T_s = 1$ s sampling time of the discrete-time LTI model of the quadruple-tank process.

Finally, Fig. 4 compares the optimal performance loss ΔJ_{opt}, the performance loss ΔJ induced by the RPI approximation $\tilde{\mathcal{Z}}_{50}^{ss}$ and the mean detection time for two different matrix M parametrizations

$$M_1 = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \text{(Fig. 4(a))}, \qquad M_2 = \begin{bmatrix} m & 0 \\ 0 & 0.5 \end{bmatrix} \text{(Fig. 4(b))}$$

and the values of m varying in the interval $m \in [1.05,\ 1.5]$. In this case, the mean detection time was obtained running 50 simulations for each value of the parameter m. The simulations show the existing trade-off between performance-loss and detection speed as a function of the maximum value ($\lambda_{\max}$) of the eigenvalues of matrix M.



(a) Matrix M_1

(b) Matrix M_2

Fig. 4. Performance loss vs mean detection time.

VIII. Conclusions

This paper has investigated the use of zonotopes for the design of watermark signals. The proposed approach has exploited the recent analogy found between stochastic and deterministic fields in the context of optimal state estimation, and extended it to optimal control. By means of this analogy, the system performance can be assessed and its steady-state operation may be related with the size of the mPRI set of the estimation error. Furthermore, similar expressions regarding the impact of a Gaussian watermark signal on the performance of an LQG system and the impact of a zonotopically bounded signal on the weighted F-norm of the states have been obtained. These analogous optimal patterns, motivate further studies in the attack detection for systems which are simultaneously subject to bounded disturbances and Gaussian noise, thus merging the usually mutually exclusive benefits granted by set-based techniques and stochastic techniques. Moreover, some unresolved issues, like the effect of the reduction operator in the convergence of the Riccati difference equations, appear as appealing problems that clearly deserve an in-depth study. Other future research directions include: the extension of the zonotopic framework to more complex attack policies like, for example, stealthy attacks; the consideration of input and state constraints; the adaptation of the watermarking method to distributed monitoring architectures, thus addressing the scalability to large-scale systems; and the consideration of continuous time systems.

Appendix A
Proof of Theorem 1

Backward induction will be used to prove (13). Given $x_N \in \langle c_N, \check{H}_N\rangle$, since its center can be rewritten as $c_N = \langle c_N, 0\rangle$, from Definition 6 it follows that $c_N^T W_f c_N = \mathcal{Q}[c_N^T W_f c_N]$. Accordingly, the expression:

$$V_N(x_N) = \mathcal{Q}[x_N^T W_f x_N] = c_N^T W_f c_N + Tr[W_f \check{P}_N]$$

satisfies (13).

Therefore, if $V_{k+1}(x_{k+1})$ satisfies (13), then, using Property 3, at a generic time instant

$$V_k(x_k) = \min_{u_k^*} \left\{ \mathcal{Q}[x_k^T W x_k] + \mathcal{Q}[u_k^T U u_k] + \mathcal{Q}[c_{k+1}^T S_{k+1} c_{k+1}] + s_{k+1} \right\}$$

with $x_k \in \langle c_k, \check{H}_k\rangle$ and $u_k \in \langle u_k, 0\rangle$.

Furthermore, from the output equation (4b), it follows that:

$$y_k \in C\langle c_k, \check{H}_k\rangle \oplus F\langle 0, I_{n_v}\rangle = \langle Cc_k, [C\check{H}_k,\ F]\rangle$$

hence, by recalling that the equation of the zonotope center (6a) generated by the ZKF is

$$c_{k+1} = (A - G_k^* C)c_k + Bu_k + G_k^* y_k$$

then the prediction zonotope center satisfies

$$c_{k+1} \in \langle (Ac_k + Bu_k), [G_k^* C\check{H}_k,\ G_k^* F]\rangle.$$

Therefore, from Definition 6, the following expressions are obtained:

$$\mathcal{Q}[x_k^T W x_k] = c_k^T W c_k + Tr[W\check{P}_k]$$

$$\mathcal{Q}[u_k^T U u_k] = u_k^T U u_k$$

$$\mathcal{Q}[c_{k+1}^T S_{k+1} c_{k+1}] = (Ac_k + Bu_k)^T S_{k+1}(Ac_k + Bu_k) + Tr[S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})]$$

with $\check{P}_k = \check{H}_k \check{H}_k^T$, $\check{H}_k = \downarrow_{q,W}(H_k)$ and $Q_v = FF^T$.

Using the expressions above, the *cost-to-go* can be rewritten as

$$V_k(x_k) = \min_{u_k^*} \left\{ u_k^{*T}(B^T S_{k+1} B + U)u_k^* + 2u_k^{*T} B^T S_{k+1} A c_k + c_k^T (W + A^T S_{k+1} A)c_k + Tr[W\check{P}_k] + Tr[S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})] + s_{k+1} \right\}$$

and thus the optimal u_k^* that minimizes $V_k(x_k)$ is

$$u_k^* = -(B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A c_k.$$

Consequently, $V_k(x_k)$ results in

$$\begin{aligned} V_k(x_k) &= c_k^T(W + A^T S_{k+1} A)c_k \\ &\quad - c_k^T(A^T S_{k+1} B(B^T S_{k+1} B + U)^{-1} B^T S_{k+1} A)c_k \\ &\quad + Tr[W\check{P}_k + S_{k+1}(G_k^*(C\check{P}_k C^T + Q_v)G_k^{*T})] + s_{k+1} \\ &= c_k^T S_k c_k + s_k. \end{aligned}$$

Since at the given time instant $x_k \in \langle c_k, \check{H}_k \rangle$, then the center satisfies $c_k = \langle c_k, 0 \rangle$ and thus $c_k^T S_k c_k = \mathcal{Q}[c_k^T S_k c_k]$, which concludes the proof. ■

Appendix B
Proof of Theorem 2

The system (20) can be rewritten as

$$e_{k+1} = \check{A} e_k + \check{B}\check{w}_k \tag{49}$$

with $\check{w}_k^T = [w_k^T,\ v_k^T] \in \check{\mathcal{W}}$ and

$$\check{A} = A - G^* C, \qquad \check{B} = [E,\ -G^* F]$$
$$\check{\mathcal{W}} = \left\langle \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_{n_w} & 0 \\ 0 & I_{n_v} \end{bmatrix} \right\rangle.$$

Besides, since $\check{\mathcal{W}}$ is a zonotopic set, and making use of (3), the following zonotope can be introduced $\bigoplus_{i=0}^{k} \check{A}^i \check{B} \check{\mathcal{W}} = \langle 0, \check{H}_{k+1} \rangle = \langle 0, [\check{A}^k \check{B},\ \check{A}^{k-1}\check{B},\ \ldots,\ \hat{B}] \rangle$, whose covariation matrix is given by $\check{P}_{k+1} = \check{H}_{k+1}\check{H}_{k+1}^T = \sum_{i=0}^{k} \check{A}^i \check{B}\check{B}^T \check{A}^{Ti}$.

Therefore, given the asymptotically stable system (49), from [36, Section 4] it is known that its mRPI set exists, its unique and it is defined as $\Phi = \bigoplus_{i=0}^{\infty} \check{A}^i \check{B} \check{\mathcal{W}}$. Accordingly, $P_\Phi = \lim_{k\to\infty} P_{k+1} = \sum_{i=0}^{\infty} \check{A}^i \check{B}\check{B}^T \check{A}^{Ti}$ is the covariation matrix of Φ, which, since $\check{A}$ is asymptotically stable, is the unique solution of Lyapunov equation $P_\Phi = \check{A} P_\Phi \check{A}^T + \check{B}\check{B}^T$ [49].

On the other hand, selecting $G^* = A P_\infty C^T (C P_\infty C^T + Q_v)^{-1}$, matrix P_∞ is the unique positive definite solution of (19), which can be rewritten in the form $P_\infty = \check{A} P_\infty \check{A}^T + \check{B}\check{B}^T$. Hence, it follows that $P_\Phi = P_\infty$, and thus $\|\Phi\|_F^2 = Tr[P_\Phi] = Tr[P_\infty]$. ■

Appendix C
Computing an Initial Zonotopic RPI Set

Below, it is presented the procedure followed in order to compute a zonotopic RPI set for a perturbed discrete-time LTI system like (2) with the disturbance set $\mathcal{W} = \langle 0, I_{n_w} \rangle$. Whenever is possible, the so-called Ultimate Bound (UB) method is used for computing a low-order initial RPI set. Under certain conditions on the closed-loop matrix A, the UB allows to directly compute a first-order zonotopic RPI set. In this regard, the following lemma presented in [50] is introduced.

Lemma 1: Suppose an invertible (complex) transformation $V \in \mathbb{C}^{n_x \times n_x}$ exists such that the matrix $\Lambda = |V^{-1} A V|$ is strictly stable. Then, the set

$$\mathcal{S} = \{x \in \mathbb{R}^{n_x} : |V^{-1}x| \le (I_{n_x} - \Lambda)^{-1} |V^{-1} B| \mathbf{1}_{n_w}\} \tag{50}$$

is an invariant set for (2) with $w_k \in \langle 0, I_{n_w} \rangle, \forall k \in \mathbb{N}$.

From (50), it follows that the resulting set $\mathcal{S}$ is a parallelotope, and thus a first-order zonotope, if V is a real matrix $V \in \mathbb{R}^{n_x \times n_x}$. Note that if matrix A has real eigenvalues, by performing the Jordan decomposition $\Lambda = V^{-1} A V$, then matrix V presents real entries. On the other hand, if A presents complex conjugate pairs of eigenvalues $\lambda_{i,j} = a \pm bi$ that satisfy (51), then the following steps can be performed in order to obtain a matrix $\bar{V}$ that satisfies Lemma 1.

1) Compute Jordan decomposition $\Lambda = V^{-1} A V$, such that Λ is an upper triangular matrix with the eigenvalues of A in the main diagonal, and V presents complex pair of columns $x \pm yi$ associated which each pair of complex conjugate eigenvalues $\lambda_{i,j} = a \pm bi$.

2) Substitute each pair of complex columns $x \pm yi$ by the pair of real vectors (x, y) that span the corresponding 2-dimensional invariant subspace of A [51]. This generates a new nonsingular matrix $\bar{V}$ such that $\bar{\Lambda} = \bar{V}^{-1} A \bar{V}$.

Note that matrix $\bar{\Lambda}$ results in an upper triangular matrix whose elements in the main diagonal are the real eigenvalues λ_r of A or 2-by-2 block matrices B_i associated with the complex conjugate pairs $\lambda_{i,j} = a \pm bi$ with the form

$$B_i = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Lemma 1 requires that the eigenvalues of $|\bar{\Lambda}|$ satisfy $|\lambda(|\bar{\Lambda}|)| < 1$. From the asymptotic stability of A it follows that the real eigenvalues λ_r already satisfy that $|\lambda_r| < 1$. On the other hand, for the complex eigenvalues it must be satisfied $|\lambda(|B_i|)| < 1$, with $\lambda(|B_i|)_{1,2} = |a| \pm |b|$. Therefore, $\bar{V}$ is a real matrix satisfying $|\lambda(|\bar{\Lambda}|)| < 1$, and thus $\mathcal{S}$ is a zonotopic RPI set, for those complex eigenvalues that satisfy

$$||a| - |b|| \le ||a| + |b|| = |a| + |b| < 1. \tag{51}$$

Finally, the parallelotope (50), can be easily transformed into a first-order zonotope in generator representation by means of the relationships presented in [52].

References

[1] X.-M. Zhang, Q.-L. Han, X. Ge, D. Ding, L. Ding, D. Yue, and C. Peng, "Networked control systems: A survey of trends and techniques," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 1, pp. 1–17, 2019.

[2] R. Langner, " Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[3] A. A. Cardenas, S. Amin, and S. Sastry, " Secure control: Towards survivable cyber-physical systems," in *Proc. 28th IEEE Int. Conf. Distributed Computing Systems Workshops*, 2008, pp. 495–500.

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

[5] 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.

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

[7] H. S. Sánchez, D. Rotondo, T. Escobet, V. Puig, and J. Quevedo, "Bibliographical review on cyber attacks from a control oriented perspective," *Annual Reviews in Control*, 2019.

[8] D. Ding, Q.-L. Han, X. Ge, and J. Wang, "Secure state estimation and control of cyber-physical systems: A survey," *IEEE Trans. Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 176–190, 2020.

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

[10] G. Franzè, F. Tedesco, and D. Famularo, " Resilience against replay attacks: A distributed model predictive control scheme for networked multi-agent systems," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 3, pp. 628–640, 2020.

[11] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. IEEE 47th Annual Allerton Conf. Communication, Control, and Computing*, 2009, pp. 911–918.

[12] Y. Mo, S. Weerakkody, and B. Sinopoli, " Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine* , vol. 35, no. 1, pp. 93–109, 2015.

[13] H. Liu, Y. Mo, and K. H. Johansson, " Active detection against replay attack: A survey on watermark design for cyber-physical systems," in *Safety, Security and Privacy for Cyber-Physical Systems*. Springer, 2021, pp. 145–171.

[14] R. M. Ferrari and A. M. Teixeira, " Detection and isolation of replay attacks through sensor watermarking," *IFAC-PapersOnLine* , vol. 50, no. 1, pp. 7363–7368, 2017.

[15] R. M. Ferrari and A. M. Teixeira, " A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks," *IEEE Trans. Automatic Control*, 2020.

[16] B. Satchidanandan and P. R. Kumar, " Dynamic watermarking: Active defense of networked cyber-physical systems," *Proc. the IEEE*, vol. 105, no. 2, pp. 219–240, 2016.

[17] H. Liu, Y. Mo, J. Yan, L. Xie, and K. H. Johansson, " An online approach to physical watermark design," *IEEE Trans. Automatic Control*, vol. 65, no. 9, pp. 3895–3902, 2020.

[18] C. Fang, Y. Qi, P. Cheng, and W. X. Zheng, "Optimal periodic watermarking schedule for replay attack detection in cyber–physical systems," *Automatica*, vol. 112, p. 108698, 2020. DOI: 10.1016/j.automatica.2019.108698.

[19] F. Miao, M. Pajic, and G. J. Pappas, " Stochastic game approach for replay attack detection," in *Proc. 52nd IEEE Conf. Decision and Control*, 2013, pp. 1854–1859.

[20] A. Khazraei, H. Kebriaei, and F. R. Salmasi, " A new watermarking approach for replay attack detection in lqg systems," in *Proc. IEEE 56th Annu. Conf. Decision and Control*, 2017, pp. 5143–5148.

[21] C. Combastel, " An extended zonotopic and Gaussian Kalman filter (EZGKF) merging set-membership and stochastic paradigms: Toward nonlinear filtering and fault detection," *Annual Reviews in Control*, vol. 42, pp. 232–243, 2016.

[22] F. Schweppe, "Recursive state estimation: Unknown but bounded errors and system inputs," *IEEE Trans. Automatic Control* , vol. 13, no. 1, pp. 22–28, 1968.

[23] V. Puig, " Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies," *Int. Journal of Applied Mathematics and Computer Science*, vol. 20, no. 4, pp. 619–635, 2010.

[24] H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu, "Set-membership estimation for complex networks subject to linear and nonlinear bounded attacks," *IEEE Trans. Neural Networks and Learning Systems*, vol. 31, no. 1, pp. 163–173, 2019.

[25] Y. Zhang, Y. Zhu, and Q. Fan, " A novel set-membership estimation approach for preserving security in networked control systems under deception attacks," *Neurocomputing*, vol. 400, pp. 440–449, 2020.

[26] C. Trapiello, V. Puig, and D. Rotondo, " A zonotopic set-invariance analysis of replay attacks affecting the supervisory layer," *Systems & Control Letters*, vol. 157, p. 105056, 2021.

[27] C. Combastel, " Zonotopes and Kalman observers: Gain optimality under distinct uncertainty paradigms and robust convergence," *Automatica*, vol. 55, pp. 265–273, 2015.

[28] D. Bertsekas, *Dynamic programming and optimal control: Volume I*. Athena scientific, 2012, vol. 1.

[29] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*. Springer Science & Business Media, 2012, vol. 1.

[30] B. Z. Temoçin and G.-W. Weber, "Optimal control of stochastic hybrid system with jumps: A numerical approximation," *Journal of Computational and Applied Mathematics*, vol. 259, pp. 443–451, 2014.

[31] E. Savku and G.-W. Weber, " A stochastic maximum principle for a Markov regime-switching jump-diffusion model with delay and an application to finance," *Journal of Optimization Theory and Applications*, vol. 179, no. 2, pp. 696–721, 2018.

[32] T. Başar and P. Bernhard, *H-Infinity Optimal Control and Related Minimax Design Problems: A Dynamic Game Approach*. Springer Science & Business Media, 2008.

[33] R. B. Vinter, "Minimax optimal control," *SIAM Journal on Control and Optimization*, vol. 44, no. 3, pp. 939–968, 2005.

[34] J. Löfberg, *Minimax Approaches to Robust Model Predictive Control*. Linköping University Electronic Press, 2003, vol. 812.

[35] C. Trapiello and V. Puig, "Replay attack detection using a zonotopic KF and LQ approach," in *Proc. IEEE Int. Conf. Systems, Man, and Cybernetics*, 2020, pp. 3117–3122.

[36] I. Kolmanovsky and E. G. Gilbert, " Theory and computation of disturbance invariant sets for discrete-time linear systems," *Mathematical Problems in Engineering*, vol. 4, no. 4, pp. 317–367, 1998.

[37] F. Blanchini and S. Miani, *Set-Theoretic Methods in Control*. Springer, 2008.

[38] C. Combastel, " A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes," in *Proc. 44th IEEE Conf. Decision and Control*, 2005, pp. 7228–7234.

[39] T. Alamo, J. M. Bravo, and E. F. Camacho, " Guaranteed state estimation by zonotopes," *Automatica* , vol. 41, no. 6, pp. 1035–1043, 2005.

[40] M. Pourasghar, V. Puig, and C. Ocampo-Martinez, " Interval observer versus set-membership approaches for fault detection in uncertain systems using zonotopes," *Int. Journal of Robust and Nonlinear Control*, vol. 29, no. 10, pp. 2819–2843, 2019.

[41] Y. Wang, V. Puig, and G. Cembrano, " Set-membership approach and Kalman observer based on zonotopes for discrete-time descriptor systems," *Automatica*, vol. 93, pp. 435–443, 2018.

[42] J. Wang, Y. Shi, M. Zhou, Y. Wang, and V. Puig, " Active fault detection based on set-membership approach for uncertain discrete-time systems," *Int. Journal of Robust and Nonlinear Control* , vol. 30, no. 14, pp. 5322–5340, 2020.

[43] R. R. Bitmead and M. Gevers, " Riccati difference and differential equations: Convergence, monotonicity and stability," in *The Riccati Equation*. Springer, 1991, pp. 263–291.

[44] D. Q. Mayne and W. Schroeder, " Robust time-optimal control of constrained linear systems," *Automatica*, vol. 33, no. 12, pp. 2103–2118, 1997.

[45] S. Olaru, J. A. De Doná, M. M. Seron, and F. Stoican, "Positive invariant sets for fault tolerant multisensor control schemes," *Int. Journal of Control*, vol. 83, no. 12, pp. 2622–2640, 2010.

[46] S. V. Rakovic, E. C. Kerrigan, K. I. Kouramas, and D. Q. Mayne, "Invariant approximations of the minimal robust positively invariant set," *IEEE Trans. Automatic Control*, vol. 50, no. 3, pp. 406–410, 2005.

[47] K. H. Johansson, " The quadruple-tank process : A multivariable laboratory process with an adjustable zero," *IEEE Trans. Control Syst. Technol.*, vol. 8, no. 3, pp. 456–465, 2000.

[48] I. ILOG, "Cplex optimizer 12.8," 2018.

[49] J. D. Hamilton, *Time Series Analysis*. Princeton university press, 2020.

[50] M. M. Seron and J. A. De Doná, " Robust fault estimation and compensation for LPV systems under actuator and sensor faults," *Automatica*, vol. 52, pp. 294–301, 2015.

[51] G. H. Golub and C. F. Van Loan, *Matrix Computations*. JHU press, 2013, vol. 3.

[52] M. Althoff, O. Stursberg, and M. Buss, "Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes," *Nonlinear Analysis: Hybrid Systems*, vol. 4, no. 2, pp. 233–249, 2010.

Carlos Trapiello received the B.Sc. degree in aerospace engineering from the Universidad Politécnica de Madrid (UPM), Spain, in 2015; and the M.Sc. degree in automatic control & robotics and the Ph.D. degree in control engineering from the Universitat Politècnica de Catalunya-BarcelonaTech (UPC), Spain, in 2018 and 2021, respectively. He is currently on a postdoctoral stay at the UMR 5218 - IMS Laboratory, France.

Vicenç Puig received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in automatic control, vision, and robotics from the Universitat Politècnica de Catalunya-BarcelonaTech (UPC), Spain, in 1993 and 1999, respectively. He is currently a Full Professor with the Automatic Control Department, UPC, and a Researcher with the Institut de Robòtica i Informàtica Industrial (IRI), CSIC-UPC. He is also the Director of the Automatic Control Department and the Head of the Research Group on Advanced Control Systems (SAC), UPC. He has developed important scientific contributions in the areas of fault diagnosis and fault-tolerant control, using interval, and linear-parameter-varying models using set-based approaches. He has participated in more than 20 European and national research projects in the last decade. He has also led many private contracts with several companies and has published more than 220 journal articles as well as over 500 contributions in international conference/workshop proceedings. He has supervised over 25 Ph.D. dissertations and over 50 master theses/final projects. He was the General Chair of the third IEEE Conference on Control and Fault-Tolerant Systems (SysTol 2016) and the IPC Chair of IFAC Safeprocess 2018. He is also the Chair of the IFAC Safeprocess TC Committee 6.4.