Detection of replay attacks in cyber-physical systems using a frequency-based signature

Helem Sabina Sánchez[a], Damiano Rotondo[a,b,*], Teresa Escobet[a], Vicenç Puig[a,b], Jordi Saludes[a], Joseba Quevedo[a]

[a]*Research Center for Supervision, Safety and Automatic Control (CS2AC), Universitat Politècnica de Catalunya (UPC), Rambla Sant Nebridi, s/n, 08022 Terrassa, Spain*
[b]*Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain*

Abstract

This paper proposes a frequency-based approach for the detection of replay attacks affecting cyber-physical systems (CPS). In particular, the method employs a sinusoidal signal with a time-varying frequency (authentication signal) into the closed-loop system and checks whether the time profile of the frequency components in the output signal are compatible with the authentication signal or not. In order to carry out this target, the couplings between inputs and outputs are eliminated using a dynamic decoupling technique based on vector fitting. In this way, a signature introduced on a specific input channel will affect only the output that is selected to be associated with that input, which is a property that can be exploited to determine which channels are being affected. A bank of band-pass filters is used to generate signals whose energies can be compared to reconstruct an estimation of the time-varying frequency profile. By matching the known frequency profile with its estimation, the detector can provide the information about whether a replay attack is being carried out or not. The design of the signal generator and the detector are thoroughly discussed, and an example based on a quadruple-tank process is used to show the application and effectiveness of the proposed method.

Keywords: Cyber-physical systems, cyber-attacks, replay attacks, signal generator, detector logic.

*Corresponding author
Email address: damiano.rotondo@yahoo.it (Damiano Rotondo)

1. Introduction

Cyber-physical systems (CPS) refer to a new generation of systems that results from the combination and coordination between the computation, communication and physical processes. This interaction through the different modalities allows developing innovative technologies, while leading to new research challenges. CPSs are ubiquitous in advanced manufacturing systems, transportation networks, industrial control processes, and critical infrastructures [1]. It is worth mentioning that the integration of cyber and physical components increases the systems' efficiency but at the same time makes them susceptible to hazards, generating in this way concerns about possible cyber-attacks targeting them.

Complex cyber-attacks capable of violating the properties of data and information technology services (confidentiality, integrity and availability [2]) have become common in recent years. Cyber-attacks compromise measurements, actuators data integrity and readiness, and have the ability of spreading within seconds. Among the most relevant cases, there are: the blackouts in large parts of Brazil, where underground railways, traffic lights, street lamps and others were all affected [3]; the Slammer worm, which in the year 2003 penetrated into the network of the Davis-Besse nuclear power plant [4], an event which created awareness in the industry about the consequences of Internet worms or virus on physical plants; and the Stuxnet malware, one of the most important attacks, which increased awareness in the public due to its complexity, functionalities and impact on the media. This malware infected industrial computer systems (compromising PLC software) and was responsible for disrupting the Iranian nuclear facility at Natanz [5]. The complexity of this attack showed that the attacker had knowledge of the data management (cyber components) and infrastructure weaknesses (physical components) of the control system.

Since cyber-attacks are generic, they can influence the physical processes through the feedback actuation, affecting many components in a coordinated way, and can be re-designed to target any other CPS. Security in control systems is not a new topic in the literature, since works about fault diagnosis and fault tolerant control techniques have been presented in [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. These approaches need to be extended to handle cyber-threats, so there is a growing interest in the study of CPS vulnerabilities and how to make these systems resilient to possible disruptions.

The effects and impact of cyber-attacks on CPS were discussed in [17], and a classification into *deception* and *denial-of-services* (DoS) attacks was provided by [18]. In particular, deception attacks consist in one or more components (sen-

sors, actuators and/or controllers) receiving false information and believing it to be true. On the other hand, DoS attacks correspond to the case in which the components cannot communicate between each other, e.g., by preventing the actuator and sensor data from reaching their destinations. The detection of DoS attacks has been investigated thoroughly during the recent years, and several approaches have been suggested, see e.g. [19, 20, 21, 22, 18, 23, 24, 25]. On the other hand, works about detection of deception attacks are more scarce in the literature since, as discussed by [26], they are subtler than DoS attacks, hence harder to detect. It is worth mentioning the solution proposed by [27], based on a state filtering scheme and sensor scheduling co-design, and [28], in which a $\mathscr{H}_\infty$ filter is designed taking into account the possibility that a neural network could be affected by deception attacks.

Replay attacks are a particular type of deception attacks. When a replay attack is carried out, at first the attacker records the measurements coming from the sensors. Then, in a subsequent phase of the attack, the attacker replaces the real data with the recorded one, causing deterioration of the control system's performance and potentially allowing to perform other types of physical attacks without being discovered. This type of attacks is often depicted in movies, where images coming from surveillance cameras are replaced with recorded videos in order to hide theft, sabotage or similar actions. The detection of these attacks was first considered by [26], where a statistical detector was employed, and a Gaussian signal (authentication signature) was added to the optimal control input in order to increase the attack detection rate, although at the cost of sacrificing the control performance. In the last few years, alternative approaches have been suggested for detecting replay attacks. For example, [29] have applied a stochastic game approach to this problem. A variation of the receding-horizon control law to deal with this kind of attacks and analyze the resulting system performance degradation was investigated by [30]. Finally, data-driven methods [31], quantized signals [32] and spectral estimation [33] are other examples of recently proposed techniques.

In this paper, we propose a new method to detect replay attacks affecting CPSs which, differently from the previously described methods, employs a frequency-based signature. This method introduces a sinusoidal signal with a time-varying frequency (authentication signal) into the closed-loop system, and checks whether the time profile of the frequency components in the output signals is compatible with the authentication signal or not. More specifically, the detection algorithm compares the energies of different signals, obtained by applying band-pass filtering to the measurements coming from the sensors. The design of the signal generator and the detector are thoroughly discussed throughout the paper, and an

example based on a quadruple-tank process is used to show the application of the method and its effectiveness in determining whether a replay attack is being carried out and, in the affirmative case, identifying which channels are being affected.

It is worth highlighting that the proposed method could be either applied alone, in situations where the applicability of other approaches could not be feasible (for example, introducing a Gaussian signature, as proposed by [26], could be problematic due to the limited bandwidth of the actuators), or it could work alongside existing methods to further enhance the capability of detecting replay attacks. Notably, when compared to other existing methods [30, 31, 32, 33], the proposed frequency-based method provides also information about which output channel is being attacked.

The rest of the paper is organized as follows. Section 2 is devoted to the problem formulation and overview of the proposed method. Then, in Section 3, the main concepts related to the signal generator are discussed. Section 4 presents the detector logic, which determines when there is a replay attack, and illustrates the choice of the design parameters. In Section 5, the proposed method is applied to an example based on a quadruple-tank process, and simulation results are used to validate its performance. Finally, the main conclusions are drawn in Section 6.

2. Problem Formulation

2.1. Description

In this section, we introduce the replay attack, which can be used by an adversary to disrupt the behavior of the system while remaining hidden.

In order to establish the scenario of a replay attack, some conditions must be taken into account:

1. The controlled system is either in a constant or a periodic steady state when the adversary performs the attack;

2. It is assumed that the attacker has control over all sensors;

3. The control loop could be disrupted because of the corrupted data.

In this work, for the sake of simplicity, linear time invariant (LTI) models are considered in order to describe the dynamic behavior of the plant. Advantages of this type of models, which will be exploited throughout the work, is that the superposition principle holds and, moreover, when excited with a sinusoidal wave input at a given frequency, the output is itself a sinusoidal wave at the same frequency,

whose magnitude and phase can be determined by looking at the frequency response characteristics. It is worth recalling that, whenever a nonlinear plant is operating around a constant steady-state, an equivalent LTI model representing the plant with a good approximation can be obtained by means of linearization.

More specifically, let us consider a continuous-time LTI system with the following state space form:

$$\dot{x}(t) = Ax(t) + Bu(t) + Dd(t) \tag{1}$$
$$y(t) = Cx(t) + Ev(t) \tag{2}$$

where $x \in \mathbb{R}^{n_x}$ is the state variable, $u \in \mathbb{R}^{n_u}$ is the action applied to the process (input), $y \in \mathbb{R}^{n_y}$ represents the sensor measurements (output), $d \in \mathbb{R}^{n_d}$ is the exogenous disturbance, $v \in \mathbb{R}^{n_v}$ represents the measurement noise, and A, B, C, D, E are known matrices of appropriate dimensions. We will assume that each element of the vectors d and v can be described by Gaussian white noise with unit variance (the case of colored noise can be taken into account by filtering white noise through a dynamic process).

Given a system, a generic adversary model applicable to an attack scenario is composed by an attack policy [34], defined as:

$$a(t) = [\tilde{u}(t), \tilde{y}(t)] = h(S, u(t), y(t)) \tag{3}$$

where $a(t)$ is the attack vector at time t, that can affect the system behavior; S represents the system knowledge including the physical plant, the controller and the detector; $u(t)$ and $y(t)$ are the available input and output data collected by the attacker; while $\tilde{u}(t)$ and $\tilde{y}(t)$ are the corrupted input and output, respectively.

Once the attack policy is defined, the replay attack can be presented. This type of attack, which does not corrupt the input $u(t)$, is carried out in two stages:

1. The attacker collects the data without disturbing the system. This stage does not affect the dynamics of the system and allows the adversary to collect knowledge that may be used in later phases of the attack. The data gathering starts from time t_0 until $t_0 + w$, where w is the size of the attack window; thus in $t \in [t_0, t_0 + w]$:
$$\begin{aligned} a(t) &= 0 \\ \widetilde{u}(t) &= u(t) \\ \widetilde{y}(t) &= y(t) \end{aligned} \tag{4}$$

2. At time t_1, the attacker begins to replay the collected data, such that the data collected in the interval $[t_0, t_0 + w]$ replaces the data in the intervals

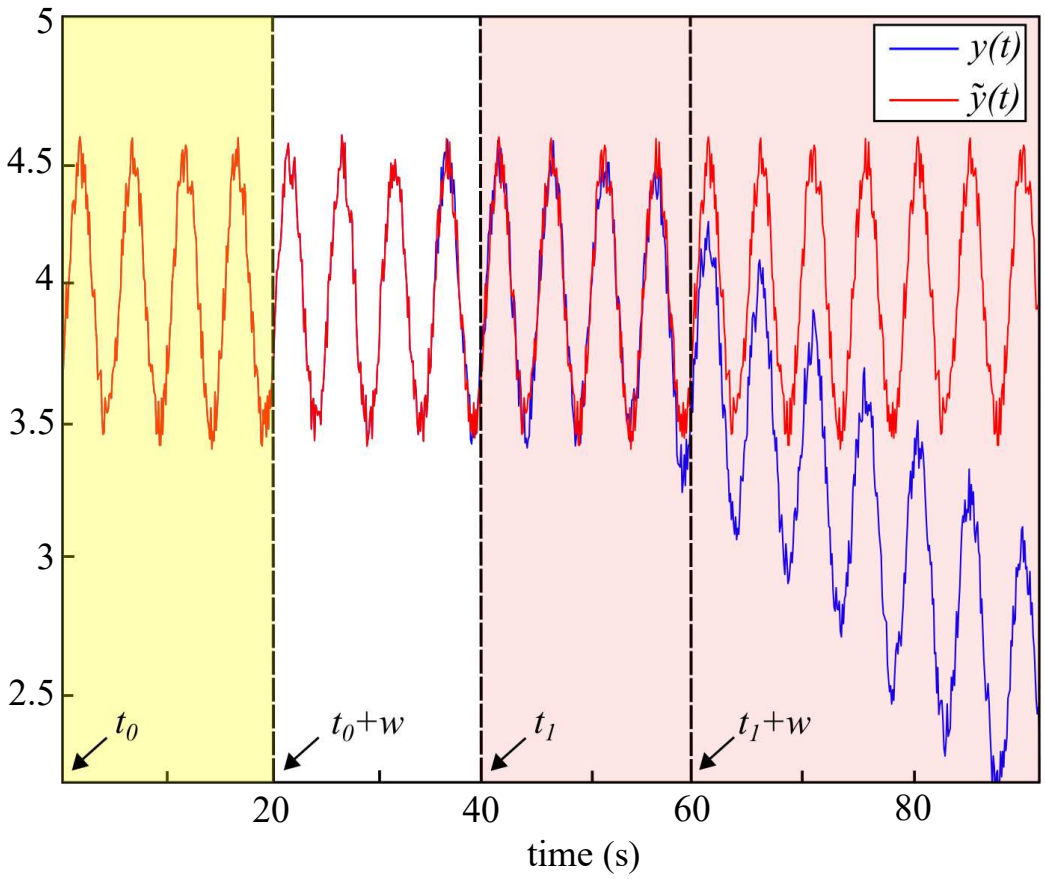

Figure 1: Replay attack example.

$\left[t_1+(N_f-1)\,w, t_1+N_f w\right]$, where $N_f \in \mathbb{N}, N_f \geq 1$:

$$\begin{aligned} a(t) &= h_f(S, y(t)) \\ \widetilde{u}(t) &= u(t) \\ \widetilde{y}(t) &= y(t+t_0-t_1-(N_f-1)w) \end{aligned} \tag{5}$$

An exemplification of a possible replay attack scenario is given in Fig. 1. In this figure, at time t_0, the attacker starts collecting the real output data $y(t)$ (blue line). Note that in this phase, the signal $\tilde{y}(t)$ (red line) matches $y(t)$ (yellow background). At time t_1, the attacker begins replaying periodically the collected data (pink background). In this particular example, it is shown that starting from time t_1+w, the attacker performs a physical attack on the system by affecting

its state, such that a mismatch between $y(t)$ and $\tilde{y}(t)$ arises (for example, this scenario could represent water or energy being stolen from a CPS). However, due to the replay attack being carried out, the physical attack goes unnoticed since the signal $\tilde{y}(t)$ is compatible with the expected system's steady state.

Remark 1. *It is noteworthy that the main goal of this attack is to make the false reading $\widetilde{y}(t)$ look as genuine as the real $y(t)$. However, as a consequence of replacing real measurements with false measurements, the feedback loop of the controller does not operate properly anymore.*

Having in mind that the control systems are not resilient to replay attacks, there is a need to develop methods to detect this kind of attack. Hereafter, the concepts behind the methodology proposed in this paper will be described.

2.2. Overview of the proposed method

The method proposed in this paper aims at detecting a replay attack by implementing a frequency-based signature method. The idea is to introduce the signature (a sinusoidal signal with a time-varying frequency) into the system and to detect if the measured output is compatible with the introduced signature or not. In order to carry out this goal, first there is a need to eliminate the couplings between inputs and outputs, such that a signature introduced on a specific input channel will affect only the output which is selected to be associated with that input. This is done through minimization of the coupling at specific frequencies, using a dynamic decoupling technique based on vector fitting. In order to carry out the detection, the output signals are passed through a bank of band-pass filters. Each filter is designed to let pass only a specific frequency among the ones used for the generation of the authentication signal. By comparing the energies of the band-pass filtered signals, an estimation $\hat{\sigma}(t)$ of $\sigma(t)$ can be determined. Then, by matching the known piecewise constant signal $\sigma(t)$ with its estimation, the detector will provide the information if a replay attack is being carried out or not.

3. Signal Generator

3.1. Description

In this section, we describe the signal generator module, which is one of the components of the frequency-based replay attack detector (see Fig. 2 for the conceptual system diagram). Following the work presented in [26], in order to detect

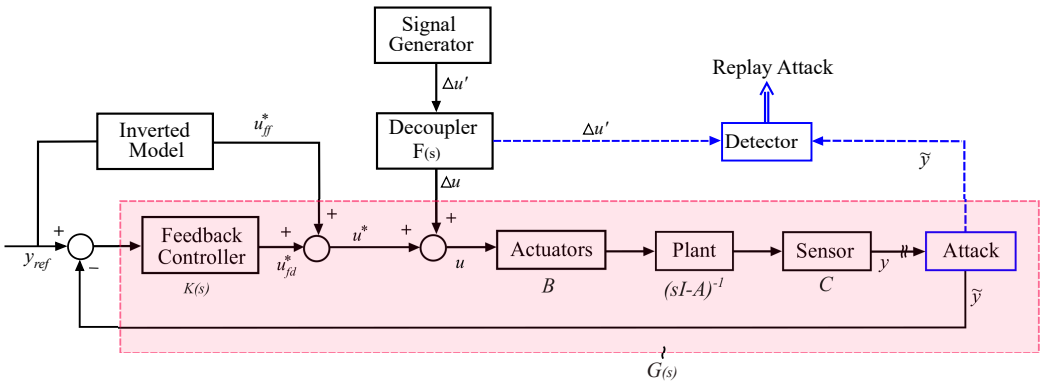

Figure 2: System Diagram.

a replay attack, it is necessary to include a *signature*, which is an authentication signal $\Delta u(t)$, into the input $u(t)$. The authors in [26] have suggested to generate the signature Δu from an independent and identically distributed Gaussian distribution with zero mean and a certain covariance, and to apply a χ^2 detector [35] to evaluate the presence of a replay attack from an anomaly in its expectation. On the other hand, the approach proposed in this paper aims at detecting replay attacks using a frequency-based signature.

In particular, $u(t)$ is made up by two different signals:

$$u(t) = u^*(t) + \Delta u(t) \tag{6}$$

where $u^*(t)$ is the control signal, chosen as the combination of a feedforward and a feedback law:

$$u^*(t) = u^*_{ff}(t) + u^*_{fb}(t) \tag{7}$$

while the signature $\Delta u(t)$ should be a zero-mean signal such that no bias is introduced in $x(t)$. Following the assumption about the controlled system being either in a constant or a periodic steady state (see Section 2), the reference trajectory $y_{ref}(t)$ can be expressed as the sum of a finite number R of sinusoidals:

$$y_{ref}(t) = \sum_{r=1}^{R} Y^{(r)}_{ref} \cos\left(\omega_r t + \varphi^{(r)}_{ref}\right) \tag{8}$$

where $Y^{(r)}_{ref} \in \mathbb{R}^{n_y}_{+}$, $\omega_r \in \mathbb{R}_{+}$ and $\varphi^{(r)}_{ref} \in \mathbb{R}^{n_y}_{[0,2\pi]}$ are the magnitude, frequency and phase, respectively, of each component ($R = 1$ and $\omega_1 = 0\,rad/s$ describe the case

of constant steady-state). As a consequence of the linearity of the system, the feedforward input $u^*_{ff}(t)$ needed to track the reference trajectory (8) is given by:

$$u^*_{ff}(t) = \sum_{r=1}^{R} U^{(r)}_{ff} \cos\left(\omega_r t + \varphi^{(r)}_{ff}\right) \tag{9}$$

with $U^{(r)}_{ff} \in \mathbb{R}^{n_u}_{+}$ and $\varphi^{(r)}_{ff} \in \mathbb{R}^{n_u}_{[0,2\pi]}$.

On the other hand, $u^*_{fb}(t)$ in (7) is a typical linear error feedback control law of the type:

$$U^*_{fb}(s) = K(s)\left(Y_{ref}(s) - Y(s)\right) \tag{10}$$

where $K(s)$ denotes the controller and $U^*_{fb}(s)$, $Y_{ref}(s)$, $Y(s)$ are the Laplace transforms of $u^*_{fb}(t)$, $y_{ref}(t)$, $y(t)$, respectively. According to the internal model principle, if the reference trajectory (8) is wanted to be tracked without steady state error, it is necessary to include its generator inside the control loop [36]. In the following, for the sake of exemplification, a constant reference trajectory will be used, such that a proportional integral (PI) structure must be chosen for $K(s)$, which can be described by [37]:

$$u^*_{fb}(t) = K_P\left(y_{ref}(t) - y(t)\right) + K_I x_I(t) \tag{11}$$

$$\dot{x}_I(t) = y_{ref}(t) - y(t) \tag{12}$$

where K_P and K_I denote the proportional and integral gain, respectively. Consequently, the system (1)-(2) can be described through the following augmented system:

$$\begin{aligned}\dot{x}_{aug}(t) = {} & A_{aug}x_{aug}(t) + \begin{bmatrix} BK_P \\ I \end{bmatrix} y_{ref}(t) \\ & + B_{aug}\left(u^*_{ff}(t) + \Delta u(t)\right) + \begin{bmatrix} D \\ 0 \end{bmatrix} d(t) + \begin{bmatrix} -BK_PE \\ -E \end{bmatrix} v(t)\end{aligned} \tag{13}$$

$$y(t) = C_{aug}x_{aug}(t) + Ev(t) \tag{14}$$

with $x_{aug}(t) = \begin{bmatrix} x(t)^T & x_I(t)^T \end{bmatrix}^T$ and:

$$A_{aug} = \begin{bmatrix} A - BK_PC & BK_I \\ -C & 0 \end{bmatrix} \qquad B_{aug} = \begin{bmatrix} B \\ 0 \end{bmatrix} \qquad C_{aug} = \begin{bmatrix} C & 0 \end{bmatrix}$$

3.2. Dynamic decoupling using vector fitting

The frequency-based signature technique aims at detecting a replay attack by introducing the authentication signal $\Delta u(t)$ into the system (1)-(2), and detecting whether the measured output is compatible with the introduced $\Delta u(t)$ or not. In order to do so, it is desirable to establish a bijection between the available inputs and the available outputs, such that the effect of an element of $\Delta u(t)$, i.e. $\Delta u_l(t)$, $l = 1,\ldots,L$, will be observed on, and only on, the associated output $y_l(t)$. However, there are two problems that hinder the establishment of such a bijection. First of all, the system (1)-(2) could be *not square*, i.e. $n_u \neq n_y$. This problem can be solved easily by considering, for replay attack detection purposes, a subset of $L = \min\{n_u, n_y\}$ inputs and outputs, such that the aforementioned bijection can be established between the elements of these subsets.

The second problem is that the closed-loop transfer matrix from $\Delta u(t)$ to $y(t)$, i.e. $G(s) = C_{aug}(sI - A_{aug})^{-1}B_{aug}$ is usually *coupled*, since each individual input influences all of the outputs. Handling these couplings (non-diagonal terms in $G(s)$) is a problem for which well-established results are available in the literature, see [38, 39, 40]. To this aim, a decoupler $F(s)$ could be introduced in the loop such that the series interconnection of $F(s)$ and $G(s)$ is *dynamically decoupled*, i.e. the transfer matrix $G_d(s) = G(s)F(s)$ is diagonal and the augmented system may be perceived as consisting of independent subsystems. However, from a practical point of view, dynamic decoupling is very demanding, since in many cases it requires a complex and highly sensitive control law, and in other cases it cannot be achieved at all [40]. For this reason, different types of *partial* decoupling have been proposed as alternatives [38], e.g. *steady-state (static) decoupling*, where a static decoupler compensates couplings at zero frequency, and *dynamic decoupling in a given frequency range*, where a dynamic decoupler minimizes couplings over a finite frequency range. However, in this paper we are interested in solving a different problem, that will be referred to as *dynamic decoupling for a given frequency set*, and which involves enforcing decoupling for a finite set of frequencies ω_i, $i = 1,\ldots,N$. The developed solution is based on *vector fitting* (VF) [41], a robust numerical method for rational approximation in the frequency domain using poles and residues.

More specifically, given the system (1)-(2), it is wished to design the decoupler:

$$\dot{x}_d(t) = A_d x_d(t) + B_d \Delta u'(t) \tag{15}$$

$$\Delta u(t) = C_d x_d(t) + D_d \Delta u'(t) \tag{16}$$

such that $\forall i = 1,\ldots,N$, $G_d(\imath\omega_i)$ calculated using $F(s) = C_d(sI - A_d)^{-1}B_d + D_d$ approximates an identity matrix. It is straightforward to obtain that, in order to achieve this goal, $F(\imath\omega_i) = G(\imath\omega_i)^{-1}$ is needed, which provides a set of N constraints that the decoupler (15)-(16) must satisfy[1].

Hence, the objective becomes approximating $F(\imath\omega_i)$, $i = 1,\ldots,N$, using a rational function, which can be chosen as [42]:

$$F(s) = \sum_{m=1}^{M} \frac{r_m}{s - a_m} + d \tag{17}$$

where M denotes the order[2], r_m and a_m are the residuals and the poles of $F(s)$, respectively, and d is a constant term.

The VF method first identifies the poles of $F(s)$ solving the following problem in the least-square sense [41, 43]:

$$\sigma(s)F(s) = p(s) \tag{18}$$

with:

$$\sigma(s) = \sum_{m=1}^{M} \frac{\tilde{r}_m}{s - q_m} + 1 \tag{19}$$

$$p(s) = \sum_{m=1}^{M} \frac{r_m}{s - q_m} + d \tag{20}$$

where $\{q_m\}$ is a set of initial poles and $\{\tilde{r}_m\}$ are the residues. The authors in [41] have shown that the poles of $F(s)$ must be equal to the zeros of $\sigma(s)$, which can be calculated as [44]:

$$\{a_m\} = \text{eig}\left(\text{diag}\{q_m\} - 1_M \cdot \tilde{r}\right) \tag{21}$$

where $\tilde{r}$ is a row vector containing $\{\tilde{r}_m\}$, and 1_M denotes a $M \times 1$ vector of ones.

The least square problem can be solved iteratively, where at each step the new poles $\{a_m\}$ replace the previous poles $\{q_m\}$ (this procedure usually converges

[1]Note that $G(\imath\omega_i)$ is a complex number, not a transfer function and, in general $F(s) \neq G(s)^{-1}$. Moreover, in cases where $G(s)$ has zeros with positive real parts, a stable $F(s)$ that satisfies $F(\imath\omega_i) = G(\imath\omega_i)^{-1}$, $i = 1,\ldots,N$, can be calculated.

[2]In general, a higher order will lead to a better approximation, but at the cost of increasing the complexity.

in 2-3 iterations). After the poles have been identified, the residues r_m can be calculated by solving once more the least square problem, this time with known poles. Finally, once $F(s)$ in (17) has been estimated, the decoupler (15)-(16) can be easily calculated, e.g. using a canonical form.

3.3. Signature generation using frequency-varying sinusoidals

The idea of the frequency-based signature approach is to introduce frequency-varying sinusoidal signals into the system (1)-(2) and the decoupler (15)-(16). The simplest possibility is to consider signals of the form:

$$\Delta u_l'(t) = \alpha_l \cos\left(\omega_{\sigma_l(t)} t\right) \qquad l = 1, \ldots, L \tag{22}$$

where α_l denotes the magnitude, while $\sigma_l(t)$ denotes a piecewise constant signal, which takes integer values between 1 and N, such that at each time instant $\omega_{\sigma_l(t)}$ equals one of the frequencies ω_i, $i = 1, \ldots, N$, for which decoupling is achieved by the decoupler (15)-(16), as explained in the previous section. It is assumed that the signal $\sigma_l(t)$ changes from its previous value to a random value between 1 and N, which could be the same as the previous value, at equally-spaced time instants $t_s^{(j)}$, $j \in \mathbb{N}_0$, with $t_s^{(0)} = 0$ and $t_s^{(j+1)} - t_s^{(j)} = T_s$, where T_s is the switching period. In the following, we will denote the value taken by $\omega_{\sigma_l(t)}$ in the time interval $[t_s^{(j)}, t_s^{(j+1)}]$ as ω_{jl}. It is worth noting that the piecewise constant signal $\sigma_l(t)$ is completely known by the detector, whereas the attacker does not have access to this information.

Let us perform the Fourier analysis of the signal (22), which can be rewritten as the sum of infinite windowed signals:

$$\Delta u_l'(t) = \alpha_l \sum_{j=0}^{\infty} w\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right) \cos\left(\omega_{jl} t\right) \tag{23}$$

where $w(\cdot)$ denotes the window function, defined as follows:

$$w\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right) = \begin{cases} 1 & t \in \left[t_s^{(j)}, t_s^{(j+1)}\right] \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Using the *linearity property* of the Fourier transform and the *convolution theorem* [45], the following is obtained from (23):

$$\Delta U_l'(\omega) = \alpha_l \sum_{j=0}^{\infty} \mathfrak{F}\left\{w\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right)\right\} * \mathfrak{F}\left\{\cos\left(\omega_{jl} t\right)\right\} \tag{25}$$

where $\mathfrak{F}\{\cdot\}$ denotes the Fourier transform of its argument, $\Delta U_l'(\omega) = \mathfrak{F}\{\Delta u_l'(t)\}$, and $*$ denotes the convolution operation. It is well-known that:

$$\mathfrak{F}\{\cos(\omega_{jl}t)\} = \pi\left[\delta(\omega+\omega_{jl}) + \delta(\omega-\omega_{jl})\right] \tag{26}$$

where $\delta(\cdot)$ is the delta function. On the other hand:

$$w\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right) = w\left(t - \frac{t_s^{(j)}+t_s^{(j+1)}}{2}, \left[-\frac{T_s}{2}, \frac{T_s}{2}\right]\right) \tag{27}$$

Hence, according to the *time shifting* property of the Fourier transform:

$$\mathfrak{F}\left\{w\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right)\right\} = e^{-\imath\omega\frac{t_s^{(j)}+t_s^{(j+1)}}{2}} W(\omega) \tag{28}$$

where [46]:

$$W(\omega) = \mathfrak{F}\left\{w\left(t, \left[-\frac{T_s}{2}, \frac{T_s}{2}\right]\right)\right\} = \frac{2\sin\left(\omega\frac{T_s}{2}\right)}{\omega} \tag{29}$$

Eq. (29) shows that the *spectral window*, i.e. the Fourier transform of the time window, decays relatively slowly (as ω^{-1}). Due to this fact, the convolution of (28) with (26) gives rise to the undesired effect known as *spectral leakage*, which was detailed carefully in the seminal work by Harris [47].

As discussed by [48], suppression of the spectral leakage can be achieved by self-convolving a window function multiple times in the time domain. More recently, [49] has presented an approach for the construction of a family of desired order continuous time window functions without self-convolution of the parent window.

Following these results, and in particular [49], another possible choice for the signal $\Delta u_l'(t)$ could be the following:

$$\Delta u_l'(t) = \alpha_l \sum_{j=0}^{\infty} w_m\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right)\cos\left(\omega_{jl}t\right) \tag{30}$$

where m is the order of the window function. For example, if $m = 1$ then $w_1\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right)$ is given by (31), which corresponds to [49]:

$$\mathfrak{F}\left\{w_1\left(t, \left[t_s^{(j)}, t_s^{(j+1)}\right]\right)\right\} = \frac{2(1-\cos\omega)}{\omega^2} e^{-\imath\omega\frac{t_s^{(j)}+t_s^{(j+1)}}{2}} \tag{32}$$

$$w_1\left(t,\left[t_s^{(j)},t_s^{(j+1)}\right]\right)=\begin{cases}1-\frac{2}{T_s}\left|t-\frac{1}{2}\left(t_s^{(j)}+t_s^{(j+1)}\right)\right| & t\in\left[t_s^{(j)},t_s^{(j+1)}\right]\\ 0 & \text{otherwise}\end{cases} \tag{31}$$

4. Detector Logic

This section describes the band-pass filtering of the output signals, the replay attack detection algorithm and the choice of the design parameters involved in the proposed strategy.

4.1. Band-pass filtering of the output signal

According to the theory of LTI systems, the response of the augmented system made up by (13)-(14) will be the sum of the natural response (which can be neglected, due to the steady-state assumption), the forced responses due to the inputs acting on it, namely $y_{ref}(t)$, $u^*_{ff}(t)$, $\Delta u'(t)$, and $d(t)$, and the noise signal $v(t)$. With the aim of analysing only the content of $y(t)$ at the frequencies ω_i, $i=1,\dots,N$, used to generate the signature signal $\Delta u'(t)$, the augmented system is cascaded with a bank of filters $H_i(s)$. In particular, each $H_i(s)$ is a $n_y\times n_y$ diagonal transfer matrix, with each element on the diagonal chosen as a band-pass filter, i.e. [50]:

$$H_i(s)=diag\left\{\frac{\frac{\omega_i}{Q_i}s}{s^2+\frac{\omega_i}{Q_i}s+\omega_i^2}\right\} \tag{33}$$

where ω_i is the frequency at which the filter peaks and Q_i is the selectivity of the filter. In general, to a higher value of Q_i corresponds a narrower frequency response $\|H_i(s)\|$ around the peak frequency ω_i, even though higher values of Q_i will also lead to a slower dynamic response, since the poles of (33) are given by:

$$s_{1/2}=-\frac{\omega_i}{2Q_i}\left(1\pm\sqrt{1-4Q_i^2}\right) \tag{34}$$

Following [51], it is possible to convert (33) into a state-space structure by using a canonical form. More specifically, by applying the observable canonical form, the l-th output of the system (1)-(2) can be fed to the following system:

$$\dot{x}_{z,il}(t)=\begin{bmatrix}0 & -\omega_i^2\\ 1 & -\omega_i/Q_i\end{bmatrix}x_{z,il}(t)+\begin{bmatrix}0\\ \omega_i/Q_i\end{bmatrix}\left(y_l(t)-y_{ref,l}(t)\right) \tag{35}$$

$$z_{il}(t)=\begin{bmatrix}0 & 1\end{bmatrix}x_{z,il}(t) \tag{36}$$

$$\hat{\sigma}_l(t) = \begin{cases} \sigma_l(t) & if\ \sigma_l(t) \neq \sigma_l(t-T_s)\ \wedge\ t \in [t_s^*, t_s^* + t_{trans} + T_\omega] \\ \arg \max\limits_{i=1,\ldots,N} \int\limits_{t-T_\omega}^{t} |z_{il}(\tau)|^2 d\tau & \text{otherwise} \end{cases} \tag{39}$$

where $y_{ref,l}(t)$ is subtracted from $y_l(t)$ in order for the band-pass filter to extract only the information that is relevant for the replay attack detection.

4.2. Replay attack detection algorithm

The replay attack detection algorithm is based on comparing the known piecewise constant signal $\sigma_l(t)$ with $\hat{\sigma}_l(t)$, which is a reconstruction based on the signals $z_{il}(t)$ obtained from (36). In particular, as long as $\hat{\sigma}_l(t) = \sigma_l(t)$, $l = 1,\ldots,L$, the algorithm will provide the information that no replay attack is being carried out on the output $y_l(t)$. On the other hand, if $\hat{\sigma}_l(t) \neq \sigma_l(t)$, then the algorithm will warn about the output $y_l(t)$ being affected by a replay attack.

It is clear that the effectiveness of the algorithm depends on how the signal $\hat{\sigma}_l(t)$ is calculated. A simple choice would be to compare the energies of the different $z_{il}(t)$ over the largest period associated with the frequencies ω_i, $i = 1,\ldots,N$, i.e. during the time intervals $[t - T_\omega, t]$, with:

$$T_\omega = \max_{i=1,\ldots,N} \frac{2\pi}{\omega_i} \tag{37}$$

and determine $\hat{\sigma}_l(t)$ as the index corresponding to the signal with the biggest energy, i.e.:

$$\hat{\sigma}_l(t) = \arg \max_{i=1,\ldots,N} \int\limits_{t-T_\omega}^{t} |z_{il}(\tau)|^2 d\tau \tag{38}$$

However, when a change in the frequency of the signal $\omega_{\sigma_l}(t)$ in (22) occurs, the system will exhibit a transient behavior with respect to the signal $\Delta u'(t)$, which will affect the matching between $\sigma_l(t)$ and $\hat{\sigma}_l(t)$. In these cases, a better choice is to take into account the time needed for such transient to become negligible, denoted in the following as t_{trans}, and calculate $\hat{\sigma}_l(t)$ as (39), where $t_s^* = \lfloor t/T_s \rfloor T_s$ denotes the last switching time.

It is worth noting that the analytical calculation of t_{trans}, although possible, is not an easy task, since the overall system made up by decoupler, plant, controller

and band-pass filter is a high order system. However, since the band-pass filters $H_i(s)$ determine the frequency content of the output signals, a reasonable estimation of t_{trans} is given by the biggest among the settling times of $H_i(s)$, $i = 1, \ldots, N$.

4.3. Choice of the design parameters

Hereafter, the choice of the design parameters involved in the proposed strategy is discussed. In particular, given the matrices A, B, C, D, E, K_P, K_I, and a reference signal $y_{ref}(t)$ as in (8), which determines univocally $u^*_{ff}(t)$, the following parameters should be determined: N, $\omega_1, \ldots, \omega_N$, $\alpha_1, \ldots, \alpha_L$, T_s and $Q_1, \ldots, Q_N$. In order to determine these parameters, the following considerations will be taken into account:

- independently from the choice of ω_i, T_s and Q_i, the gain from the signal $\Delta u'(t)$ to the output of the band-pass filter (33) will be an identity matrix and, in order for the attacker not to realize about the presence of $\Delta u'(t)$ by looking at the output signal coming from the sensors, $\Delta u'(t)$ should be *small* when compared to $y_{ref}(t)$;

- in order for the attacker not to realize about the presence of $\Delta u'(t)$ by looking at the input signals being sent to the actuators, $\Delta u(t)$ should be *small* when compared to $u^*_{ff}(t)$;

- $\Delta u'(t)$ should overcome the effect of the unknown disturbance $d(t)$ and the measurement noise $v(t)$ on the output, denoted in the following as $y_d(t)$ and $y_v(t)$, respectively;

- the filter selectivities Q_i should be chosen such that the components of $y(t)$ at frequencies $\omega_j \neq \omega_i$ are attenuated sufficiently; however, Q_i cannot be too high, because such a choice would lead to a slower response of the band-pass filter, as shown by Eq. (34);

- higher frequencies ω_i are desirable in order to make the response of the band-pass filters faster; however, ω_i cannot be too high because typically stronger $\Delta u'(t)$ are needed at high frequencies in order to overcome the effect of the measurement noise, due to the limited band of the actuators;

- the switching period T_s should be big enough such that the outputs of the band-pass filters settle to the corresponding steady-state after a change in the frequency $\omega_{\sigma_l}(t)$ in (22);

- finally, the number of frequencies N is a degree of freedom in the design of the detector, which should be selected in order to obtain faster replay attack detectors.

A first constraint on $\Delta u'(t)$ aims at making this signal *small* when compared to $y_{ref}(t)$:

$$A_{peak}\left\{\Delta u'_l(t)\right\} \ll A_{peak}\left\{y_{ref,l}(t)\right\} \qquad l = 1,\ldots,L \tag{40}$$

where $A_{peak}\{\cdot\}$ denotes the peak amplitude. It is straightforward that $A_{peak}\left\{\Delta u'_l(t)\right\} = \alpha_l$ while, on the other hand:

$$A_{peak}\left\{y_{ref,l}(t)\right\} \leq \sum_{r=1}^{R} Y^{(r)}_{ref,l} \tag{41}$$

which means that (40) can be rewritten as:

$$\alpha_l < \kappa_1 \sum_{r=1}^{R} Y^{(r)}_{ref,l} \qquad l = 1,\ldots,L \tag{42}$$

with $\kappa_1 \ll 1$.

Another constraint on $\Delta u'_l(t)$ aims at making $\Delta u(t)$ small when compared to $u^*_{ff}(t)$:

$$A_{peak}\left\{\Delta u_l(t)\right\} \ll A_{peak}\left\{u^*_{ff,l}(t)\right\} \qquad l = 1,\ldots,L \tag{43}$$

In order to estimate $A_{peak}\left\{\Delta u_l(t)\right\}$, let us note first that, independently from the choice of ω_i, $F(\imath\omega_i) = G(\imath\omega_i)^{-1}$ will hold by design. Then, by neglecting the spectral leakage, the following relationship can be obtained:

$$A_{peak}\left\{\Delta u_l(t)\right\} \leq \sum_{m=1}^{L} \max_{i=1,\ldots,N} |F_{lm}(\imath\omega_i)|\,\pi\alpha_m \tag{44}$$

while, on the other hand:

$$A_{peak}\left\{u^*_{ff,l}(t)\right\} \leq \sum_{r=1}^{R} U^{(r)}_{ff,l} \tag{45}$$

which means that (43) can be rewritten as:

$$\sum_{m=1}^{L} \max_{i=1,\ldots,N} |F_{lm}(\imath\omega_i)|\,\pi\alpha_m < \kappa_2 \sum_{r=1}^{R} U^{(r)}_{ff,l} \quad l = 1,\ldots,L \tag{46}$$

with $\kappa_2 \ll 1$.

With regard to the effect of the unknown disturbance $d(t)$ and the measurement noise $v(t)$ on the output, simple calculations show that the transfer functions from d and v to y are given by:

$$T_{yd}(s) = \left[I + C(sI - A)^{-1}BK(s)\right]^{-1} C(sI - A)^{-1}D \tag{47}$$

$$T_{yv}(s) = \left[I + C(sI - A)^{-1}BK(s)\right]^{-1} E \tag{48}$$

Following [52], and taking into account that both d and v are independent white noises, i.e. their power spectral densities are identity matrices, the power spectral density of $y_d(t) + y_v(t)$ can be calculated as:

$$S_y(\omega) = \left[\begin{array}{cc} \overline{T_{yd}}(j\omega) & \overline{T_{yv}}(j\omega) \end{array}\right] \left[\begin{array}{c} T_{yd}(j\omega)^T \\ T_{yv}(j\omega)^T \end{array}\right] \tag{49}$$

where the bar denotes the conjugate operation. Then, a possible specification concerning $\Delta u'(t)$ overcoming the effect of the unknown signals can be expressed as:

$$A_{peak}\left\{\Delta u'_l(t)\right\} = \alpha_l > \kappa_3 \sigma_l \qquad l = 1, \ldots, L \tag{50}$$

with $\kappa_3 \gg 1$, where σ_l is the standard deviation of the l-th element of $y_d(t) + y_v(t)$, which can be calculated from the l-th diagonal element of $S_y(\omega)$, namely $S_{y,ll}(\omega)$, as follows:

$$\sigma_l = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{y,ll}(\omega) d\omega} \tag{51}$$

In order for each filter $H_i(s)$ to reject adequately the frequency content corresponding to values ω_j of the varying frequency which are different from the specific ω_i of the filter, it is suggested to choose the ω_i sufficiently spaced among themselves. For example, by requiring that $|H_i(j\omega_{i-1})| \leq \Psi$ and $|H_i(j\omega_{i+1})| \leq \Psi$, conditions (52)-(53) are obtained.

In fact, from (33), it follows that:

$$|H_i(\imath\omega)| = \frac{\omega_i \omega}{\sqrt{Q_i^2\left(\omega_i^2 - \omega^2\right)^2 + \omega_i^2 \omega^2}} \tag{54}$$

By requiring that $|H_i(\imath\omega)| = \Psi$, the following equation is obtained:

$$\Psi^2 Q_i^2 \omega^4 + \left[\left(\Psi^2 - 1\right) - 2Q_i^2\Psi^2\right] \omega_i^2 \omega^2 + Q_i^2 \Psi^2 \omega_i^4 = 0 \tag{55}$$

$$\omega_{i-1} \leq \frac{\omega_i}{\Psi Q_i} \sqrt{\frac{2\Psi^2 Q_i^2 + (1-\Psi^2) - \sqrt{(\Psi^2-1)^2 - 4\Psi^2(\Psi^2-1)Q_i^2}}{2}} \tag{52}$$

$$\omega_{i+1} \geq \frac{\omega_i}{\Psi Q_i} \sqrt{\frac{2\Psi^2 Q_i^2 + (1-\Psi^2) + \sqrt{(\Psi^2-1)^2 - 4\Psi^2(\Psi^2-1)Q_i^2}}{2}} \tag{53}$$

---

that has the solution:

$$\omega = \frac{\omega_i^2 \left(2\Psi^2 Q_i^2 + \left(1-\Psi^2\right) \pm \sqrt{(\Psi^2-1)^2 - 4\Psi^2(\Psi^2-1)Q_i^2} \right)}{2\Psi^2 Q_i^2} \tag{56}$$

which leads to (52)-(53).

However, it can be calculated that, if equal rejection properties are desired for the frequencies ω_{i-1} and ω_{i+1}, then Q_i should satisfy:

$$Q_i = \frac{\omega_{i-1}\omega_i\sqrt{1-\Psi^2}}{\Psi\left|\omega_i^2 - \omega_{i-1}^2\right|} = \frac{\omega_{i+1}\omega_i\sqrt{1-\Psi^2}}{\Psi\left(\omega_{i+1}^2 - \omega_i^2\right)} \tag{57}$$

which means that $\omega_{i-1} = \omega_i/k$ and $\omega_{i+1} = k\omega_i$ for some $k > 1$, i.e. all the frequencies should be selected as elements of a geometric series. Under this choice, it can be shown that $Q_i = Q$, $i = 1,\ldots,N$, with:

$$Q = \frac{k\sqrt{1-\Psi^2}}{\Psi(k^2-1)} \tag{58}$$

Concerning the choice of the switching period T_s, taking into account the discussion in Section 4.2 and the reconstruction of $\hat{\sigma}_l(t)$ using (39), it is clear that the following should hold:

$$T_s \gg t_{trans} \tag{59}$$

which leads to:

$$T_s = \kappa_4 t_{trans} \tag{60}$$

with $\kappa_4 \gg 1$.

Hence, (42), (46), (50), (58) and (60) provide conditions for a suitable choice of the design parameters, in the form of a set of inequalities to be solved under

the constraint that $\omega_i = k^{i-1}\omega_1$, $i = 1,\ldots,N$. The frequency ω_1 can be chosen as the highest frequency for which a solution to the set of inequalities can be found. On the other hand, it is worth noting that (46) leads to a tradeoff between the number of different frequencies ω_i that can be used, i.e. the design parameter N, and how high the frequency ω_1 can be chosen. Hence, the choice of N must take into account that, on one hand, the phenomenon of coincidental matches between the generated and the reconstructed random frequency profiles should be avoided and, on the other hand, the band-pass filters should be faster in order to obtain shorter settling times, which would allow for a smaller T_s, ultimately leading to a faster replay attack detector.

5. Example

In this section, the signal generator and the detector logic presented in the previous sections are illustrated by considering a quadruple-tank process controlled through a wireless communication network (see Fig. 3), which is a testbed that has found recent success in the field of secure control against cyber attacks [53].

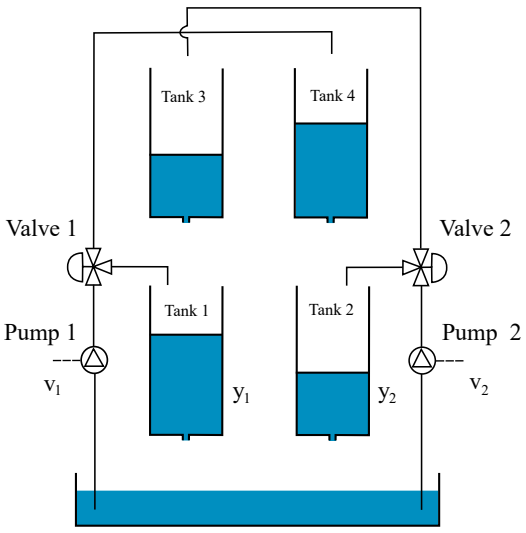

Figure 3: Schematic diagram of the quadruple-tank process.

Table 1: Parameters of the quadruple-tank process

Symbol	Description	Parameter value
A_1, A_3	Cross-sections of Tanks 1,3	$28\,cm^2$
A_2, A_4	Cross-sections of Tanks 2,4	$32\,cm^2$
a_1, a_3	Cross-sections of outlet holes 1,3	$0.071\,cm^2$
a_2, a_4	Cross-sections of outlet holes 2,4	$0.057\,cm^2$
g	Acceleration of gravity	$981\,cm/s^2$
k_1	Flow parameter from tank 1 to 4	$3.14\,cm^3/Vs$
k_2	Flow parameter from tank 2 to 3	$3.29\,cm^3/Vs$
γ_1	Valve 1 opening parameter	0.43
γ_2	Valve 2 opening parameter	0.34

The plant model is given by [54]:

$$\begin{aligned}
\frac{dh_1(t)}{dt} &= -\frac{a_1}{A_1}\sqrt{2gh_1(t)} + \frac{a_3}{A_1}\sqrt{2gh_3(t)} + \frac{\gamma_1 k_1}{A_1}u_1(t) \\
\frac{dh_2(t)}{dt} &= -\frac{a_2}{A_2}\sqrt{2gh_2(t)} + \frac{a_4}{A_2}\sqrt{2gh_4(t)} + \frac{\gamma_2 k_2}{A_2}u_2(t) \\
\frac{dh_3(t)}{dt} &= -\frac{a_3}{A_3}\sqrt{2gh_3(t)} + \frac{(1-\gamma_2)k_2}{A_3}u_2(t) \\
\frac{dh_4(t)}{dt} &= -\frac{a_4}{A_4}\sqrt{2gh_4(t)} + \frac{(1-\gamma_1)k_1}{A_4}u_1(t)
\end{aligned}$$

where $h_i \in [0,30]cm$, $i = 1,2,3,4$, are the state variables, corresponding to the water levels in each tank, while u_1, u_2 are the control inputs, i.e., the voltages applied to pump 1 and pump 2. The list and values of the model's parameters are given in Table 1.

The linearized state-space model around an equilibrium point is described by (1)-(2), with:

$$A = \begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{A_3}{A_1 T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{A_4}{A_2 T_4} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix} \quad B = \begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix}$$

where the time constants T_i are:

$$T_i = \frac{A_i}{a_i}\sqrt{\frac{2h_i^0}{g}} \quad i = 1,2,3,4$$

and h_i^0, $i = 1,2,3,4$ are the steady-state levels.

In particular, we will consider that the quadruple-tank system operates around the set-point $y_{ref}(t) = [6.3, 6.5]^T$, which corresponds to the feedforward actions $u_{ff}^1(t) = u_{ff}^2(t) = 3.15V$ such that $A_{peak}\left\{u_{ff}^*(t)\right\} = \left[\begin{array}{cc} 3.15 & 3.15 \end{array}\right]^T$ and steady-state equilibrium levels $h_1^0 = 12.4\,cm$, $h_2^0 = 13.2\,cm$, $h_3^0 = 4.7\,cm$, $h_4^0 = 5.0\,cm$. Consequently, the values of the matrices A and B to be used in the remaining of the example are given as follows:

$$A = \left[\begin{array}{cccc} -0.0159 & 0 & 0.0258 & 0 \\ 0 & -0.0109 & 0 & 0.0177 \\ 0 & 0 & -0.0258 & 0 \\ 0 & 0 & 0 & -0.0177 \end{array}\right] \quad B = \left[\begin{array}{cc} 0.0482 & 0 \\ 0 & 0.0350 \\ 0 & 0.0775 \\ 0.0559 & 0 \end{array}\right]$$

By considering that the measured level signals are $y_1(t) = 0.5h_1(t)$, and $y_2(t) = 0.5h_2(t)$, and that a matrix E is used to describe the sensor noise, the LTI state-space description is completed[3] by the matrices:

$$C = \left[\begin{array}{cccc} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \end{array}\right] \quad E = \left[\begin{array}{cc} 0.01 & 0 \\ 0 & 0.01 \end{array}\right]$$

In order to track $y_{ref}(t)$ with zero steady-state error, a decentralized linear error feedback PI control law as in (10), with parameters taken from [54], has been used:

$$K(s) = \left[\begin{array}{cc} \frac{165s+1.5}{110s} & 0 \\ 0 & \frac{-26.4s-0.12}{220s} \end{array}\right]$$

On the other hand, using (51), the values for the standard deviations are calculated as $\sigma_1 = 0.0032$ and $\sigma_2 = 0.0030$.

By solving inequalities (42), (46) and (50) with $\kappa_1 = \kappa_2 = \kappa_3 = 3$, and by selecting $N = 2$ and $\omega_2 = 2\omega_1$, the following parameters are calculated: $\omega_1 = 0.30\,rad/s$, $\alpha_1 = 0.0125$, $\alpha_2 = 0.0093$. Then, by requiring an attenuation of $-20\,dB$ ($\Psi = 0.1$) at frequencies ω_{i-1} and ω_{i+1}, (58) can be used to calculate Q as $Q = 2\sqrt{11}$.

According to Section 3.2, the specification of dynamic decoupling for the frequencies ω_1 and ω_2 is satisfied if $F(s)$ is chosen such that:

$$F(\imath\omega_1) = \left[\begin{array}{cc} 2.175 + 12.238\imath & -2.336 - 0.112\imath \\ -1.157 - 0.005\imath & 0.525 + 16.947\imath \end{array}\right]$$

[3]Note that no exogenous disturbance affects the plant, i.e. $D = 0$.

$$F(\imath\omega_2) = \begin{bmatrix} 2.164+24.782\imath & -2.366-0.058\imath \\ -1.169-0.003\imath & 0.508+34.218\imath \end{bmatrix}$$

Using the VFIT3 routine[4], which is an implementation of fast relaxed VF [41, 42, 55], the decoupler (15)-(16) which guarantees the above specification is calculated as:

$$A_d = \begin{bmatrix} -1061.1 & 0 & 0 & 0 \\ 0 & -0.1 & 0 & 0 \\ 0 & 0 & -0.3 & 0 \\ 0 & 0 & 0 & -1016.5 \end{bmatrix} \quad B_d = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$C_d = 10^7 \begin{bmatrix} -4.64 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5.88 \end{bmatrix} \quad D_d = 10^4 \begin{bmatrix} 4.372 & 0 \\ 0 & 5.786 \end{bmatrix}$$

A comparison between the Bode plot of the non-decoupled system (blue line) and the decoupled one (red line) is depicted in Fig. 4. It can be seen that, at the frequencies ω_1 and ω_2, G_d approximates an identity matrix such that a good decoupling is achieved.

Then, following the discussion in Section 4.3, the t_{trans} is calculated as $t_{trans} = 174\,s$ and, by applying (60) with $\kappa_4 = 4$, $T_s = 696\,s$ is obtained.

In order to assess the effectiveness of the proposed strategy, three different simulation scenarios are considered.

5.1. Scenario 1

In the first scenario, the system is working without replay attacks being performed. Fig. 5 shows the output signal $y(t)$, which tracks the reference $y_{ref}(t)$, in scenario 1. It can be seen that the introduction of the signature $\Delta u'(t)$ does not have a visible effect on $y(t)$, which is important for the attacker not to become aware of the implementation of the proposed detection strategy.

In Fig. 6, the outputs of the band-pass filters $z_{il}(t)$, $i = 1,2$, $l = 1,2$, are plotted along with the signals $\sigma_1(t)$ and $\sigma_2(t)$, which determine the time-varying frequency profile of the signal (22). It appears evident that when $\omega_{\sigma_l} = \omega_1$ (low state of the red line), then $z_{1l}(t)$ is the signal with the strongest energy. Conversely, when $\omega_{\sigma_l} = \omega_2$ (high state of the red line), then $z_{2l}(t)$ becomes the signal with the strongest energy.

Using (39), $\hat{\sigma}_1(t)$ and $\hat{\sigma}_2(t)$ can be determined, as shown in Fig. 7, and these estimations can be compared with the signals $\sigma_1(t)$ and $\sigma_2(t)$ in order to obtain a

[4]https://www.sintef.no/projectweb/vectfit/

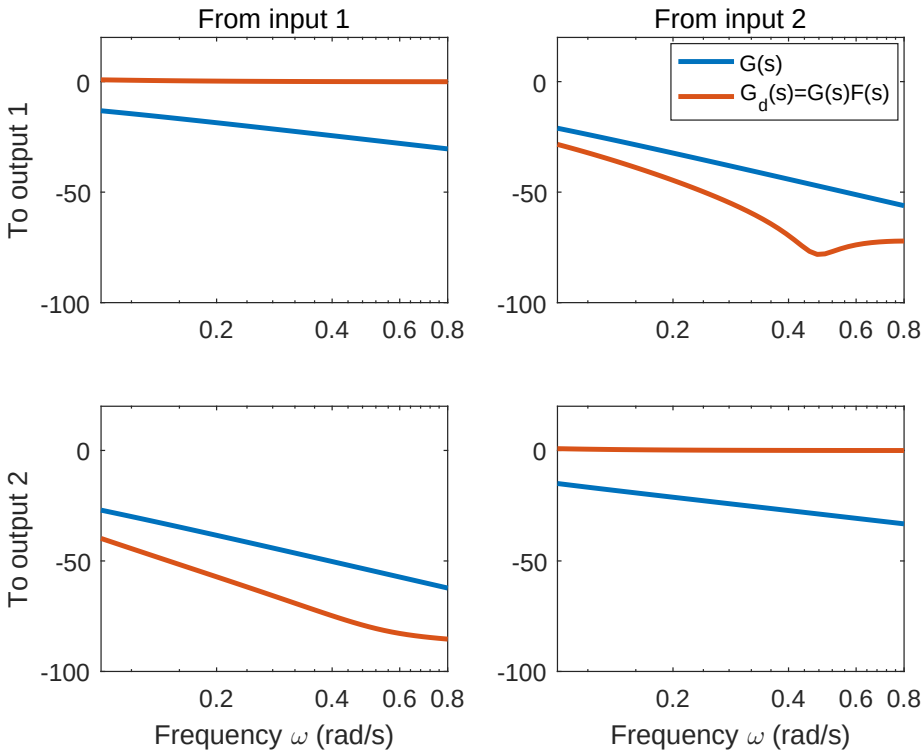

Figure 4: Decoupling (Bode plot).

Boolean information about the presence of a replay attack, as depicted in Fig. 8. It can be seen that the detection test is not affected by false alarms, hence it succeeds in providing the information about no replay attacks affecting the measurements.

5.2. Scenario 2

In the second simulation scenario, it is assumed that an attacker records the measurements of both outputs in the first $2000\,s$ and then replays the recorded data periodically starting from $t = 2000\,s$. In this case, the signals $z_{il}(t)$ do not follow anymore the corresponding varying frequency profiles $\omega_{\sigma_l(t)}$ (see Fig. 9). This fact leads to a mismatch between $\sigma_l(t)$ and $\hat{\sigma}_l(t)$, as shown in Fig. 10, which provides an information about both the output channels being attacked (see Fig. 11). In fact, based on the information provided by $\hat{\sigma}_l(t)$, a replay attack acting on the first output channel is detected at time $t = 2031\,s$, while a replay attack on the second output channel is detected at time $t = 2044\,s$ (notice that due to $\hat{\sigma}_l(t)$ being

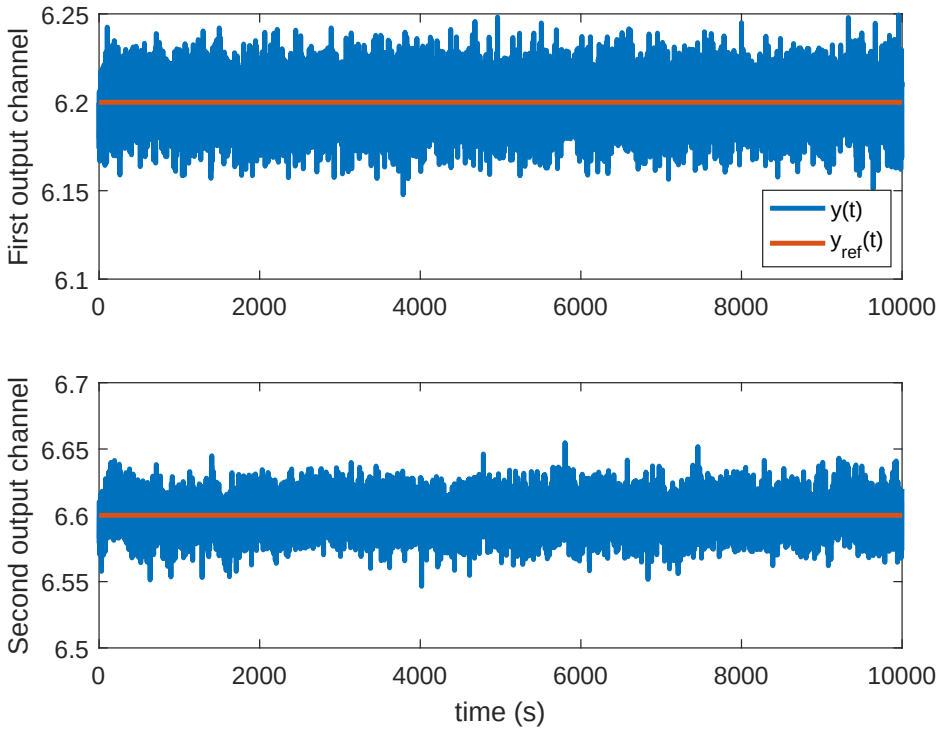

Figure 5: Output signal $y(t)$ and reference $y_{ref}(t)$ in scenario 1.

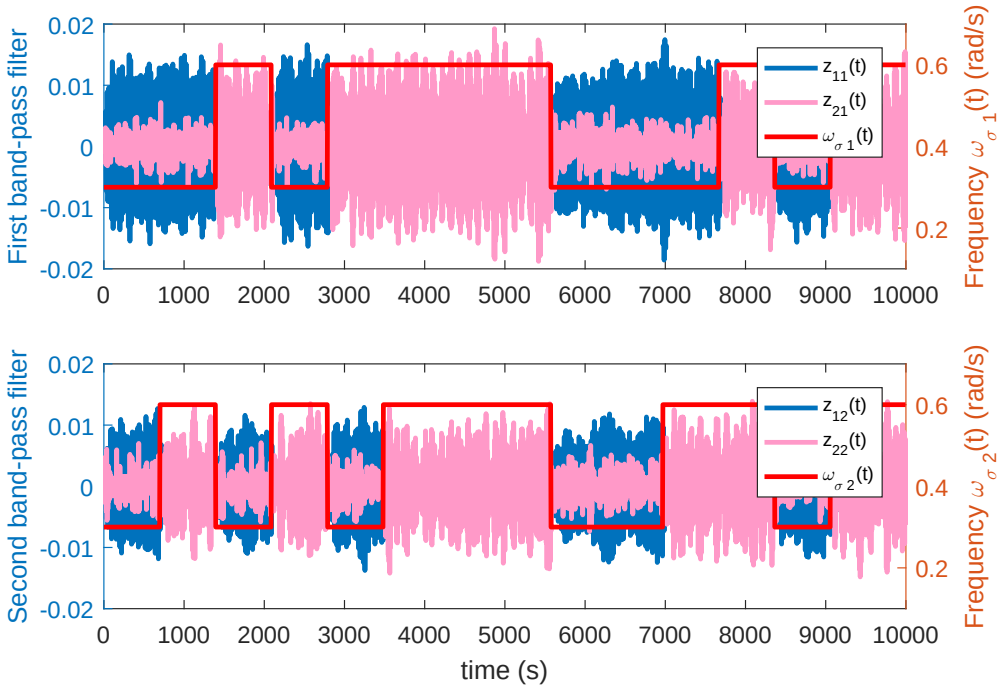

Figure 6: Outputs of the band-pass filters $z_{il}(t)$ and varying frequency $\omega_{\sigma}(t)$ in scenario 1.

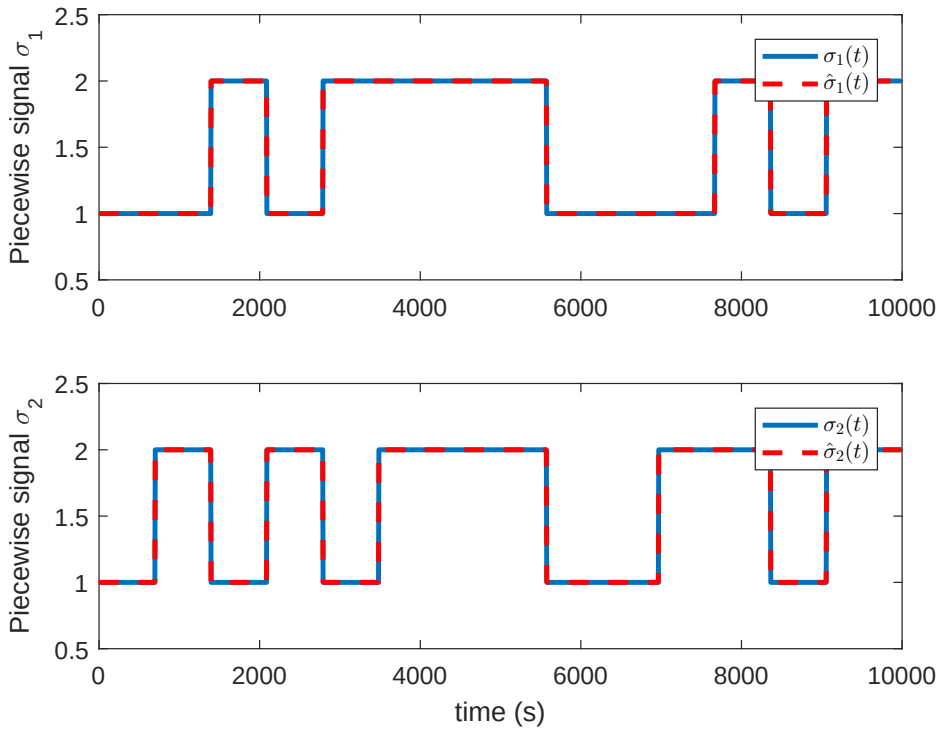

Figure 7: Piecewise constant signals $\sigma_1(t)$, $\sigma_2(t)$ and their estimations $\hat{\sigma}_1(t)$, $\hat{\sigma}_2(t)$ in scenario 1.

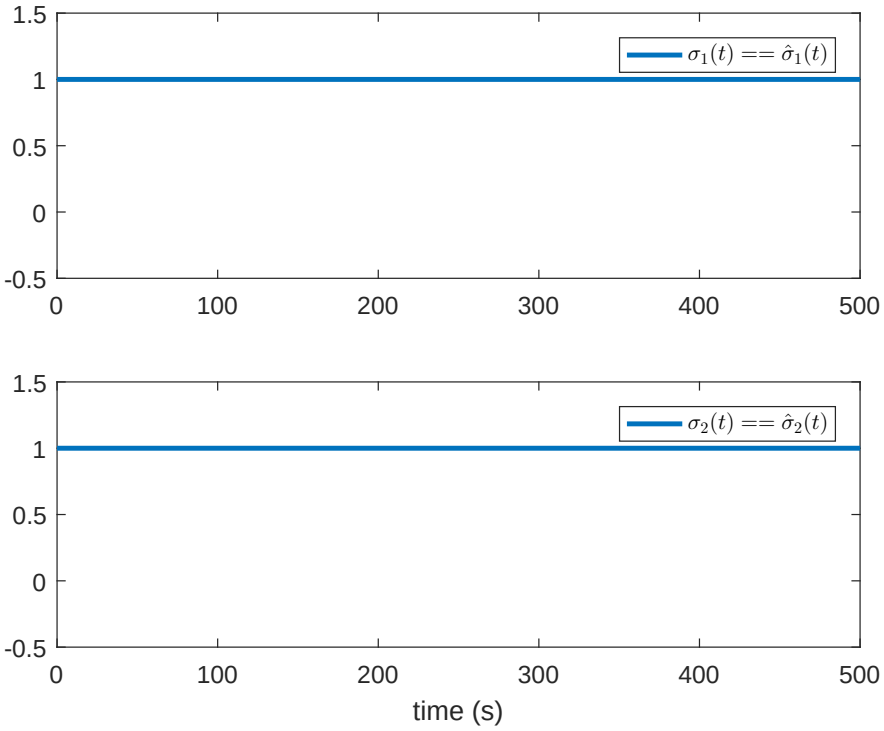

Figure 8: Result of the replay attack detection test in scenario 1.

calculated as in (39), $\sigma_1(t) \neq \hat{\sigma}_1(t)$ and $\sigma_2(t) \neq \hat{\sigma}_2(t)$ hold intermittently under replay attack).

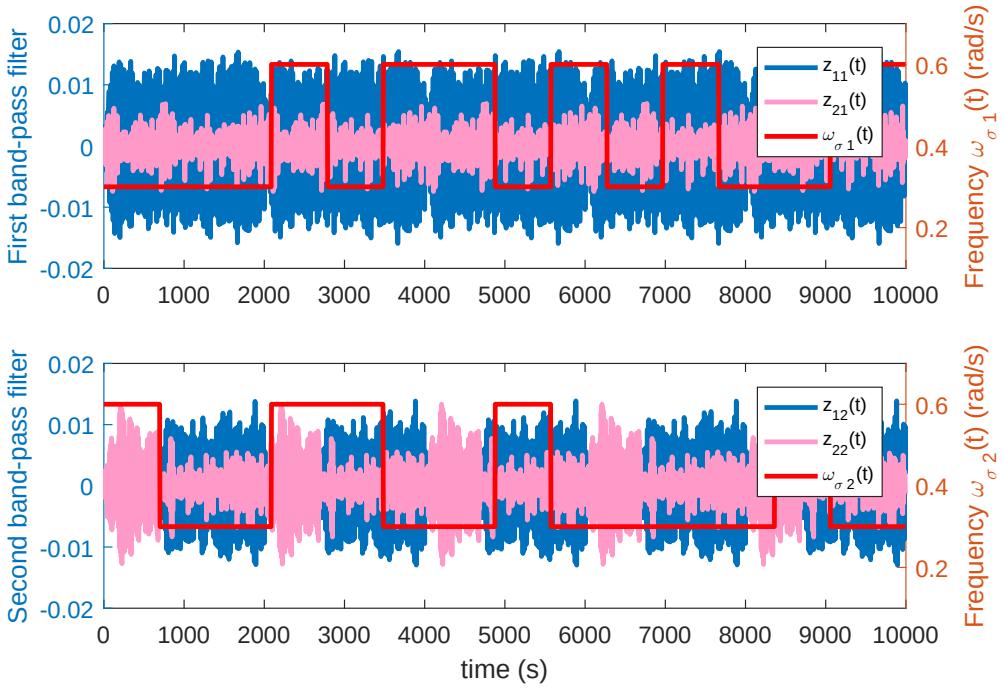

Figure 9: Outputs of the band-pass filters $z_{il}(t)$ and varying frequency $\omega_{\sigma}(t)$ in scenario 2.

5.3. Scenario 3

In the last scenario, only the first output is considered to be affected by the replay attack starting from $t = 200\,s$. In this case, the signals $z_{i1}(t)$, $i = 1,2$, do not follow the profile of $\omega_{\sigma_1}(t)$, while the signals $z_{i2}(t)$, $i = 1,2$, follow $\omega_{\sigma_2}(t)$ throughout the simulation (see Fig. 12). Consequently, a mismatch between $\sigma_1(t)$ and $\hat{\sigma}_1(t)$ arises, as shown in Fig. 13, which allows detecting a replay attack acting on the first output channel at time $t = 2032\,s$ (see Fig. 14).

5.4. Comparison between $N = 2$ and $N = 4$

In order to conclude the analysis of the performance of the proposed approach, a comparison between the detector designed for a number of frequencies $N = 2$

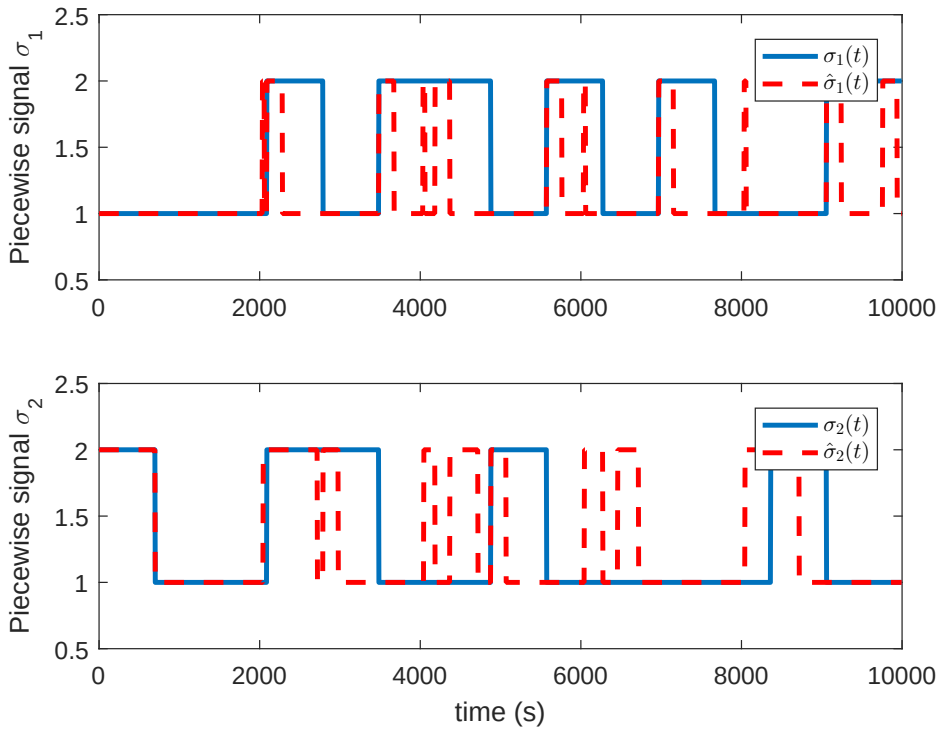

Figure 10: Piecewise constant signals $\sigma_1(t)$, $\sigma_2(t)$ and their estimations $\hat{\sigma}_1(t)$, $\hat{\sigma}_2(t)$ in scenario 2.

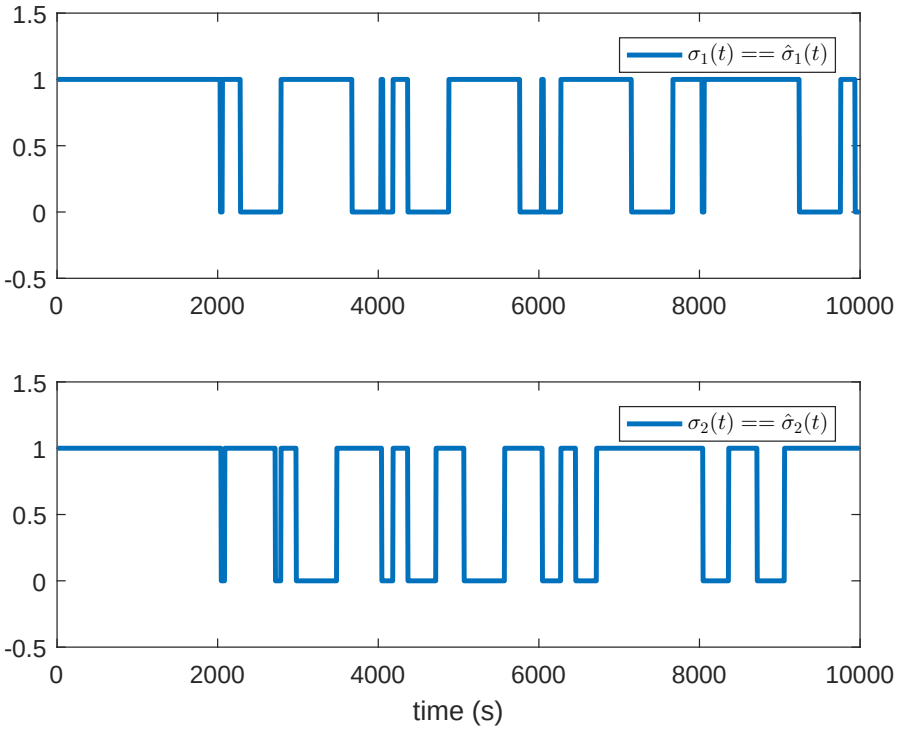

Figure 11: Result of the replay attack detection test in scenario 2.

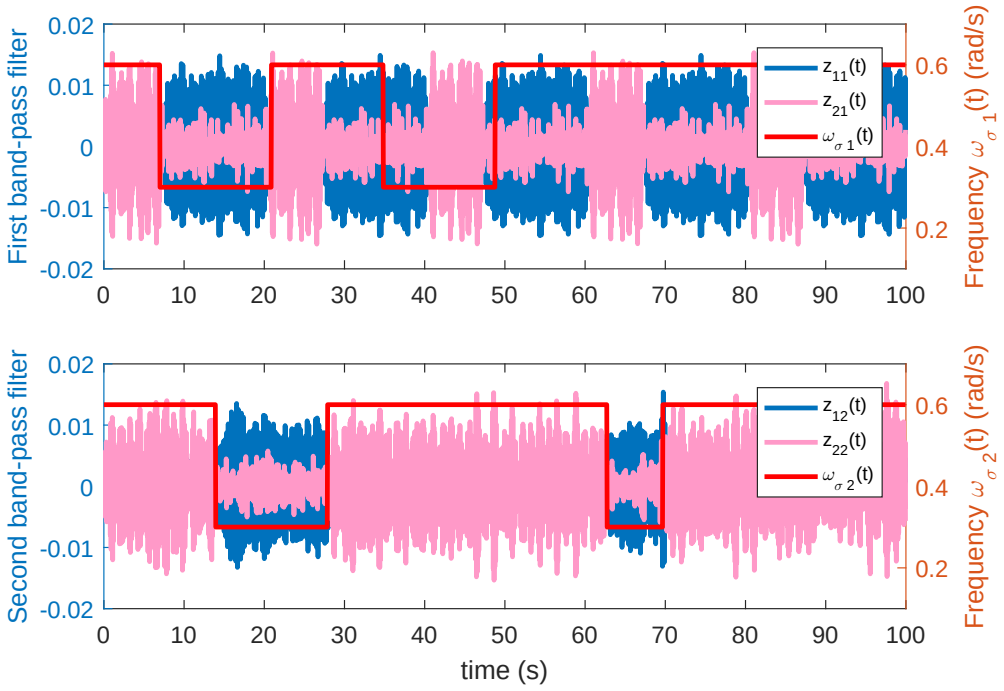

Figure 12: Outputs of the band-pass filters $z_{il}(t)$ and varying frequency $\omega_{\sigma}(t)$ in scenario 3.

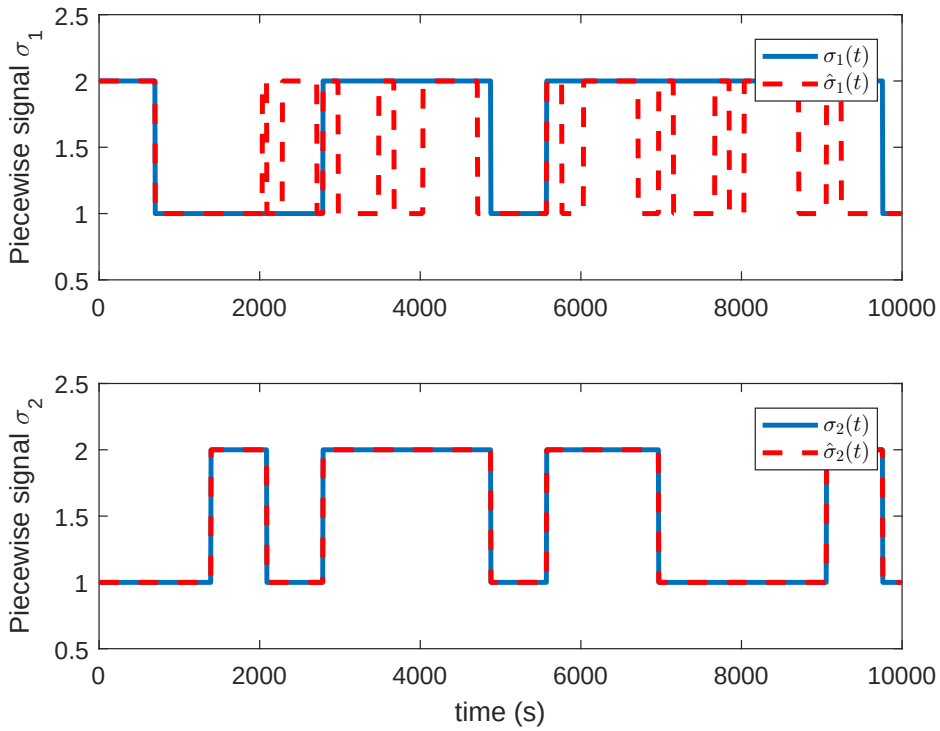

Figure 13: Piecewise constant signals $\sigma_1(t)$, $\sigma_2(t)$ and their estimations $\hat{\sigma}_1(t)$, $\hat{\sigma}_2(t)$ in scenario 3.

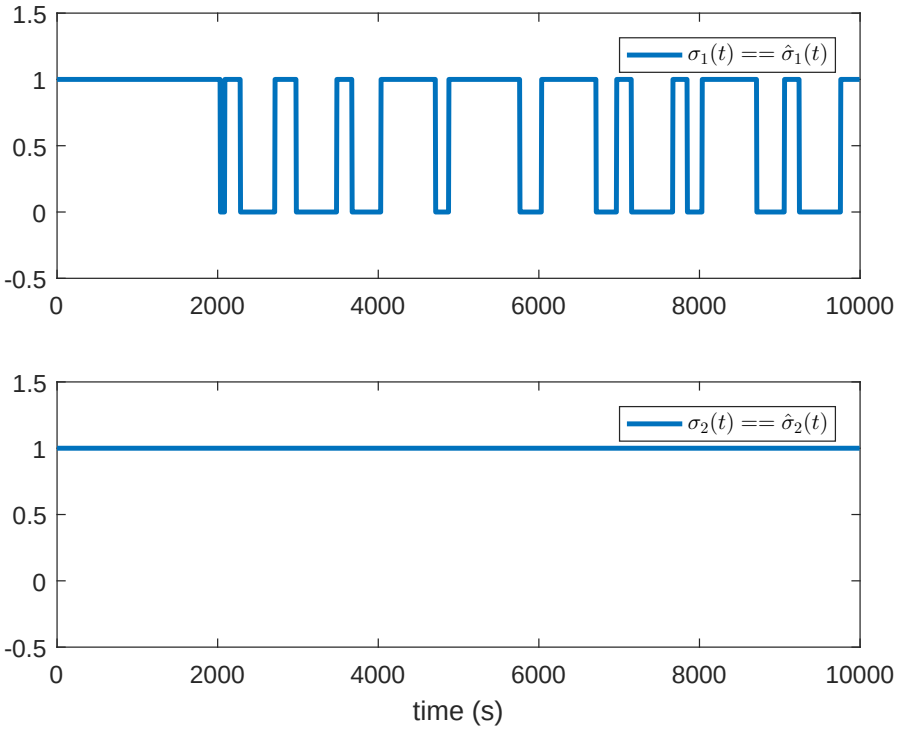

Figure 14: Result of the replay attack detection test in scenario 3.

and $N = 4$ is performed. Note that the case $N = 2$ corresponds to the same designed parameters used in the simulations of scenarios 1-3 described previously. On the other hand, the choice $N = 4$ leads to $\omega_1 = 1.4\,rad/s$, $t_{trans} = 37.40\,s$ and $T_s = 149.60\,s$ (see Fig. 15 for an exemplification of the outputs of the band-pass filters for this detector when no replay attack is affecting the system throughout a simulation). For each case, 100 simulations have been performed, in each of which a replay attack affected both of the output channels starting from a time t_1 randomly generated from a uniform distribution with support $[100\,s, 300\,s]$. Over the considered simulations, the detector with $N = 2$ has detected a replay attack acting on the first (second) output channel in an average time of $400\,s$ ($458\,s$), while the detector with $N = 4$ has performed the detection in an average time of $2694\,s$ ($2605\,s$). This comparison suggests that choosing a binary set of frequencies leads to a better performance of the detector.

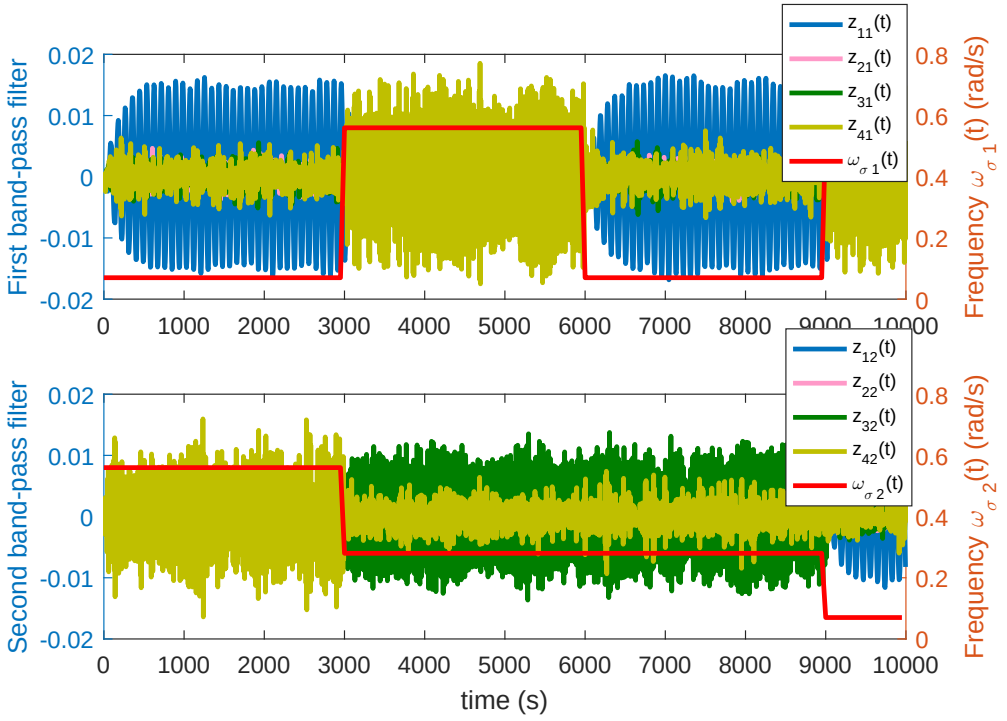

Figure 15: Outputs of the band-pass filters $z_{il}(t)$ and varying frequency $\omega_\sigma(t)$ with $N = 4$ (no replay attacks).

6. Conclusions

In this work, *replay attacks* on cyber-physical systems were considered, and an innovative method for detecting this type of attacks affecting control systems has been proposed. The developed approach is based on adding an authentication signal (signature), to the input. In particular, the chosen signature is frequency-based, which means that frequency-varying sinusoidal signals are used. More specifically, a piecewise constant signal $\sigma(t)$ determines at each instant of time the frequency of the authentication signal. By filtering the signature with a dynamic decoupler, designed using the *vector fitting* method, it is ensured that a given signature affects only one of the available output channels. This property can be exploited in order to determine which channels are being affected by the replay attack. By filtering the output signals using a bank of band-pass filters, each one designed to let pass only the component corresponding to a specific frequency among the ones used for the generation of the authentication signal, an estimation $\hat{\sigma}(t)$ of $\sigma(t)$ can be determined. Then, by comparing the known piecewise constant signal $\sigma(t)$ with its estimation, an information about whether a replay attack is being carried out or not is provided (replay attack detection algorithm). The choice of the design parameters involved in the proposed strategy has been discussed thoroughly. Finally, the signal generator and the detector logic have been evaluated by considering an example based on a quadruple-tank process. Three simulation scenarios have demonstrated the effectiveness of the proposed technique, and shown its main characteristics. In particular, the proposed method has shown not to trigger false alarms while being able to identify successfully the channels affected by the replay attack in all the considered scenarios. The comparison between detector designed with different numbers of frequencies ($N = 2$ and $N = 4$) has suggested that choosing a binary set of frequencies leads to a better performance of the detector.

Future work will aim at extending the proposed approach to discrete-time systems, as well as to add more complexity to the problem formulation by taking into account possible nonlinearity and structural uncertainties affecting the system's matrices. Moreover, the information provided by the proposed detection method will be used to develop secure control strategies, with the aim of compensating the negative effects of replay attacks, which can potentially disrupt a control system when wrong measurements are fed back to the observer/controller instead of the true ones.

7. Acknowledgements

This work has been partially funded by the Spanish State Research Agency (AEI) and the European Regional Development Fund (ERFD) through the projects DEOCS (ref. MINECO DPI2016-76493) and SCAV (ref. MINECO DPI2017-88403-R), by AGAUR of the Generalitat de Catalunya through the Advanced Control Systems (SAC) group grant (2017 SGR 482) and by the AEI through the Maria de Maeztu Seal of Excellence to IRI (MDM-2016-0656) and the grant Juan de la Cierva -Formacion (FJCI-2016-29019). The authors would like to thank Massimiliano Rotondo for his valuable comments and discussions about the application of frequency-varying signals.

References

[1] F. Pasqualetti, F. Dörfler, F. Bullo, Attack detection and identification in cyber-physical systems, IEEE Transactions on Automatic Control 58 (2013) 2715–2729.

[2] M. Bishop, Computer Security: Art and Science, Addison-Wesley Professional, 2002.

[3] J. P. Conti, The day the samba stopped, Engineering & Technology 5 (2010) 46–47.

[4] D. Moore, V. Paxson, S. Savage, C. Shannon, S. Staniford, N. Weaver, Inside the Slammer worm, IEEE Security & Privacy 99 (2003) 33–39.

[5] J. P. Farwell, R. Rohozinski, Stuxnet and the future of cyber war, Survival 53 (2011) 23–40.

[6] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, J. Schröder, Diagnosis and fault-tolerant control, Springer Berlin, 2006.

[7] I. Hwang, S. Kim, Y. Kim, C. E. Seah, A survey of fault detection, isolation, and reconfiguration methods, IEEE Transactions on Control Systems Technology 18 (2010) 636–653.

[8] K. Ji, Y. Lu, L. Liao, Z. Song, D. Wei, Prognostics enabled resilient control for model-based building automation systems, in: Proceedings of the 12th Conference of International Building Performance Simulation Association (2011), pp. 286–293.

[9] C. G. Rieger, K. Villez, Resilient control system execution agent (ReCoSEA), in: Proceedings of the 5th IEEE International Symposium on Resilient Control Systems (2012), pp. 143–148.

[10] W. C. Lin, K. Villez, H. E. Garcia, Experimental validation of a resilient monitoring and control system, Journal of Process Control 24 (2014) 621–639.

[11] D. Rotondo, V. Puig, F. Nejjari, J. Romera, A fault-hiding approach for the switching quasi-LPV fault tolerant control of a four-wheeled omnidirectional mobile robot, IEEE Transactions on Industrial Electronics 62 (2015) 3932–3944.

[12] D. Rotondo, F. R. López-Estrada, F. Nejjari, J. C. Ponsart, D. Theilliol, V. Puig, Actuator multiplicative fault estimation in discrete-time LPV systems using switched observers, Journal of the Franklin Institute 353 (2016) 3176–3191.

[13] K. Zhou, J. C. Doyle, K. Glover., Robust and optimal control, Prentice-Hall, Inc., Upper Saddle River, NJ, USA., 1996.

[14] M. Rodrigues, H. Hamdi, N. B. Braiek, D. Theilliol, Observer-based fault tolerant control design for a class of LPV descriptor systems, Journal of the Franklin Institute 351 (2014) 3104 – 3125.

[15] D. Rotondo, Advances in gain-scheduling and fault tolerant control techniques, Springer, 2017.

[16] F. Gomez-Bravo, J. M. García, R. J. Naharro, J. G. Galán, M. S. Raya, Experimental platform for studying hardware vulnerabilities on mobile robots: I2c bus, a case of study, Revista Iberoamericana de Automática e Informática Industrial RIAI 14 (2017) 205–216.

[17] A. A. Cárdenas, S. Amin, S. Sastry, Research challenges for the security of control systems, in: Proceedings of the 3rd Conference on Hot topics in security (2008), p. 6.

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

[19] G. Carl, G. Kesidis, R. R. Brooks, S. Rai, Denial-of-service attack-detection techniques, IEEE Internet computing 10 (2006) 82–89.

[20] Z. Tan, A. Jamdagni, X. He, P. Nanda, R. P. Liu, Denial-of-service attack detection based on multivariate correlation analysis, in: Proceedings of the International Conference on Neural Information Processing (2011), pp. 756–765.

[21] L. Zhao, G.-H. Yang, Adaptive sliding mode fault tolerant control for nonlinearly chaotic systems against DoS attack and network faults, Journal of the Franklin Institute 354 (2017) 6520–6535.

[22] Y. Xiang, K. Li, W. Zhou, Low-rate DDoS attacks detection and traceback by using new information metrics, IEEE Transactions on Information Forensics and Security 6 (2011) 426–437.

[23] A. Gupta, C. Langbort, T. Başar, Optimal control in the presence of an intelligent jammer with limited actions, in: Proceedings of the 49th IEEE Conference on Decision and Control (2010), pp. 1096–1101.

[24] R. Zhang, P. Venkitasubramaniam, Stealthy control signal attacks in linear quadratic gaussian control systems: Detectability reward tradeoff, IEEE Transactions on Information Forensics and Security 12 (2017) 1555–1570.

[25] B. Li, R. Lu, W. Wang, K.-K. R. Choo, DDOA: A dirichlet-based detection scheme for opportunistic attacks in smart grid cyber-physical system, IEEE Transactions on Information Forensics and Security 11 (2016) 2415–2425.

[26] Y. Mo, B. Sinopoli, Secure control against replay attacks, in: Proceedings of the 47th IEEE Annual Allerton Conference on Communication, Control, and Computing (2009), pp. 911–918.

[27] M. A. Sid, S. Chitraganti, K. Chabir, Medium access scheduling for input reconstruction under deception attacks, Journal of the Franklin Institute 354 (2017) 3678–3689.

[28] J. Liu, J. Xia, E. Tian, S. Fei, Hybrid-driven-based h_∞ filter design for neural networks subject to deception attacks, Applied Mathematics and Computation 320 (2018) 158–174.

[29] F. Miao, M. Pajic, G. J. Pappas, Stochastic game approach for replay attack detection, in: Proceedings of the 52nd IEEE Conference on Decision and Control (2013), pp. 1854–1859.

[30] M. Zhu, S. Martínez, On the performance analysis of resilient networked control systems under replay attacks, IEEE Transactions on Automatic Control 59 (2014) 804–808.

[31] M. Ma, P. Zhou, D. Du, C. Peng, M. Fei, H. M. AlBuflasa, Detecting replay attacks in power systems: A data-driven approach, in: Advanced Computational Methods in Energy, Power, Electric Vehicles, and Their Integration, Springer, 2017, pp. 450–457.

[32] K. Kashima, D. Inoue, Replay attack detection in control systems with quantized signals, in: Proceedings of the IEEE European Control Conference (2015), pp. 782–787.

[33] B. Tang, L. D. Alvergue, G. Gu, Secure networked control systems against replay attacks without injecting authentication noise, in: Proceedings of the IEEE American Control Conference (2015), pp. 6028–6033.

[34] A. Teixeira, D. Pérez, H. Sandberg, K. H. Johansson, Attack models and scenarios for networked control systems, in: Proceedings of the 1st ACM International Conference on High Confidence Networked Systems (2012), pp. 55–64.

[35] R. Mehra, J. Peschon, An innovations approach to fault detection and diagnosis in dynamic systems, Automatica 7 (1971) 637–660.

[36] R. Costa-Castelló, J. M. Olm, H. Vargas, G. A. Ramos, An educational approach to the internal model principle for periodic signals, International Journal of Innovative Computing, Information and Control 8 (2012) 5591–5606.

[37] G. F. Franklin, J. D. Powell, M. L. Workman, Digital Control of Dynamic Systems, Addison Wesley Longman, 3rd. edition, 1997.

[38] S. Skogestad, I. Postlethwaite, Multivariable Feedback Control: Analysis and Design, Wiley, 2005.

[39] O. N. Gasparyan, Linear and nonlinear multivariable feedback control: a classical approach, John Wiley and Sons, Ltd., 2008.

[40] Q.-G. Wang, Decoupling Control, Lecture Notes in Control and Information Sciences, Vol. 285, Springer-Verlag Berlin Heidelberg, 2003.

[41] B. Gustavsen, A. Semlyen, Rational approximation of frequency domain responses by vector fitting, IEEE Transactions on Power Delivery 14 (1999) 1052–1061.

[42] B. Gustavsen, Improving the pole relocating properties of vector fitting, IEEE Transactions on Power Delivery 21 (2006) 1587–1592.

[43] A. Semlyen, B. Gustavsen, Vector fitting by pole relocation for the state equation approximation of nonrational transfer matrices, Circuits, Systems and Signal Processing 19 (2000) 549–566.

[44] B. Gustavsen, A. Semlyen, Simulation of transmission line transient using vector fitting and modal decomposition, IEEE Transactions on Power Delivery 13 (1998) 605–614.

[45] R. N. Bracewell, The Fourier transform and its applications, Boston: McGraw-Hill, 2000.

[46] L. J. Van Vliet, Windowed Fourier transform, Lectures about Signals and Systems (2002).

[47] F. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, Proceedings of the IEEE 66 (1978) 51–83.

[48] N. Geckinli, D. Yavuz, Some novel windows and a concise tutorial comparison of window families, IEEE Transactions on Acoustics, Speech, and Signal Processing 26 (1978) 501–507.

[49] P. Singla, T. Singh, Desired order continuous polynomial time window functions for harmonic analysis, IEEE Transactions on Instrumentation and Measurement 59 (2010) 2475–2481.

[50] H. Zumbahlen, Linear circuit design handbook, Elsevier Newnes Press, 2008.

[51] K. Ogata, Modern control engineering, Prentice Hall, 1970.

[52] A. Le Bot, Foundation of statistical energy analysis, Oxford University Press, 2015.

[53] A. Teixeira, I. Shames, H. Sandberg, K. H. Johansson, A secure control framework for resource-limited adversaries, Automatica 51 (2015) 135–148.

[54] K. H. Johansson, The quadruple-tank process: A multivariable laboratory process with an adjustable zero, IEEE Transactions on control systems technology 8 (2000) 456–465.

[55] D. Deschrijver, M. Mrozowski, T. Dhaene, D. De Zutter, Macromodeling of multiport systems using a fast implementation of the vector fitting method, IEEE Microwave and Wireless Components Letters 18 (2008) 383–385.