Adaptive control-theoretic detection of integrity attacks against cyber-physical industrial systems

Jose Rubio-Hernan, Luca de Cicco, Joaquin Garcia-Alfaro

Adaptive control-theoretic detection of integrity attacks against cyber-physical industrial systems

Jose Rubio-Hernan[1] | Luca De Cicco[2] | Joaquin Garcia-Alfaro[1]

[1]SAMOVAR, Telecom SudParis, CNRS, Université Paris-Saclay, Evry, France

[2]Politecnico di Bari, Dipartimento di Ingegneria Elettrica e dell'Informazione, Bari, Italy

Correspondence
Joaquin Garcia-Alfaro, SAMOVAR, Telecom SudParis, CNRS, Université Paris-Saclay, Evry, France.
Email: joaquin.garcia_alfaro@telecom-sudparis.com

Funding information
Cyber CNI Chair of the Institut Mines-Télécom; Airbus Defence and Space, Amossys, EDF, Orange, La Poste, Nokia, Société Générale and the Regional Council of Brittany

Abstract

The use of control-theoretic solutions to detect attacks against cyber-physical industrial systems is a growing area of research. Traditional literature proposes the use of control strategies to retain, eg, satisfactory close-loop performance, as well as safety properties, when a communication network connects the distributed components of a physical system (eg, sensors, actuators, and controllers). However, the adaptation of these strategies to handle security incidents is an ongoing challenge. In this paper, we survey the advantages of a watermark-based detector against some integrity attacks as well as the weaknesses against other attacks. To cover these weaknesses, we propose a new control and security strategy that complements the watermark-based detector. We validate the detection efficiency of the new strategy via numeric simulation. Experimental results are also presented by using a laboratory testbed based on supervisory control and data acquisition industrial protocols.

1 | INTRODUCTION

In an effort to reducing complexity and costs, industrial infrastructures are being upgraded with novel computing, communication, and interconnection capabilities. The goal is to build and operate them as easily and reliably as possible. Current trends, such as smart factories and intelligent supply chains mainly rely on sensors, actuators, and controllers that are networked to communicate with each other by using cyber-physical industrial protocols.[1,2] However, this opens challenging security threats.[3] The recently coined cyber-physical security term refers to mechanisms that address security issues associated to these environments. The use of inadequate cyber-physical security mechanisms can have an adverse effect in smart industries.[4-6] These issues place the study of cyber-physical security mechanisms as a hot research topic.

Given the control-theoretic nature of cyber-physical industrial systems, the control community is actively working to adapt traditional control strategies to detect faults and errors, toward detectors of malicious attacks.[7-9] Motivated by the same objectives, we present in this paper a solution that combines 2 different control strategies to handle integrity attacks against cyber-physical systems. The new solution combines an event-triggered control strategy together with a challenge-response watermark-based detector. We present the theoretical approach underlying our proposed strategy as well as its algorithmic implementation. We validate our proposed approach via numerical simulations, complemented by experimental results obtained using a laboratory testbed based on supervisory control and data acquisition (SCADA) industrial protocols.

This paper is organized as follows. Section 3 provides introductory background. Section 4 presents our detection strategy. Section 5 validates the approach via numerical simulations. Section 6 complements the simulations with experimental results obtained with a laboratory testbed based on SCADA protocols. Section 2 surveys related work. Section 7 concludes this paper.

2 | RELATED WORK

Security of cyber-physical systems is drawing a great deal of attention recently.[10] Solutions focused on control approaches for the detection of cyber-physical attacks are the research axis more closely related to this paper. This axis is the one that explicitly considers the interconnection between cyber and physical control domains in networked control systems. Recently, the control system community started to study the security of cyber-physical systems both under the methodological point of view and from a more technological standpoint by looking at particular problems arising in, eg, smart grids. Concerning the methodological aspects, several studies have proposed to adapt classical frameworks to handle security issues in networked control systems.

In a cyber-physical replay attack, the adversary aims at carrying out malicious actions without holding *a priori* knowledge about system model dynamics. Most of the other cyber-physical attacks reported in the literature require that the adversary gathers some system knowledge. For example, to execute a dynamic false-data injection attack,[11] the adversary has to have a perfect knowledge of the plant's behavior. Likewise, to execute a covert attack,[12] it is necessary to have some knowledge about the model of the plant and the controller. The adversaries handled in this paper fit those aforementioned models defined in the works of Garone et al,[11] and Smith,[12] ie, it is assumed that they aim at obtaining knowledge about the model underlying the behavior of the plant and the controllers prior attacking the system.

Concerning detection mechanisms, previous work in the related literature considers the adaptation of fault detection systems to detect cyber-physical attacks.[13-15] In particular, Mo et al[13] and Mo and Sinopoli[16] proposed the use of the watermark-based detection by adapting traditional failure detection mechanisms to detect replay attacks. This watermark is added to the control measurements to verify using the detection mechanism that the sensor measurements are not replayed measurements, ie, the control measurements with the watermark have to be correlated with the sensor measurements. Miao et al[17] improved the performance of this detection mechanism. They present a suboptimal algorithm against replay attack, using a stochastic game approach. This suboptimal algorithm combines the watermark-based detector proposed by Mo et al and the zero-sum stochastic game proposed by Zhu and Başar.[18] In the same way, Do et al[19] formulated the attack detection problem as a transient changes detection problem in stochastic-dynamical systems. This detector is based on the knowledge of the behavior of the system and its stochastic variations to detect data manipulation. Using this detection control algorithm allows to protect the system against attacks able to perform malicious actions in a short period of time, of which goal is to disrupt safety-critical applications. An alternative approach is taken by Pasqualetti et al,[15] where the authors propose to use the theory of geometric control to model cyber-physical systems attacks. In line with the aforementioned approach, this paper focuses on the interconnection between control strategies and a watermark detector to handle the integrity attacks.

3 | BACKGROUND

3.1 | Cyber-physical attacks

Some of the abbreviations used hereinafter are summarized in Table 1. The coined cyber-physical industrial system term integrates a physical infrastructure and a cyber framework in an effort to automate physical processes, as well as reducing complexity and costs of traditional industrial systems. These traditional systems are composed of sensors, actuators, and other devices that

TABLE 1 Abbreviations used in this paper

Acronyms	Descriptions
CPS	Cyber-physical system
SCADA	Supervisory control and data acquisition
NCS	Networked control system
IT	Information technology
ETC	Event-triggered control
PETC	Periodical event-triggered control
LTI	Linear time invariant
LQG	Linear quadratic Gaussian
PIETC-WD	Periodical and intermittent event-triggered control and watermark detector
PLC	Programmable logic controller
RTU	Remote terminal units
ARX	Autoregressive with exogenous input

interact with the physical processes. The technology evolution brings these systems toward a combination between a physical layer, which encompasses the physical framework, and a cyber layer that encompasses the communication and computation framework,[20] eg, by using SCADA technologies.

Cyber-physical systems combine control strategies through the communication network that are used to manage the physical layer with some cyber strategies and used to monitor the physical layer. The cyber layer has to follow some constraints driven by control strategies that also appear as constraints at the physical layer. Attacks targeting both layers are often referred to as cyber-physical attacks. In a nutshell, cyber-physical attacks aim at affecting physical processes of a networked control system by manipulating network information at the cyber layer.

Teixeira et al[14] proposed a taxonomy of cyber-physical attacks based on the resources of the adversaries. For instance, the knowledge of the adversary about the system is an important resource to assess the nature of the attacks. Based on the degree of the adversary knowledge, the attacks may succeed at violating some system properties, such as availability and integrity, as well as to obtaining operational information about the system itself to cover the attacks from detection. Mo et al[13] addressed the traditional replay attacks, in which the adversary does not hold any *a priori* knowledge about the system dynamics. They also address injection attacks, as a variant of the previous type, in which the adversary injects false data or alterations from legitimate data.

Those aforementioned attacks are also addressed by Teixeira et al[14] but classified in a more granular manner (depending on the techniques to inject the malicious data) as surge attacks, bias injection attacks, and geometric attacks. In order to perform such attacks, the adversary must hold some partial knowledge about the physical processes underlying the target system. Smith[12] presented a more powerful attack type, referred to as covert attacks, in which the adversary must perfectly know the system dynamics, to evade detection. To acquire the system knowledge, adversaries are assumed to learn and identify as much information as possible from the network or physical data. Rubio-Hernan et al[21,22] differentiated between parametric and nonparametric adversaries, depending on the identification techniques used to obtain such system knowledge.

In terms of attack detection, several techniques have been proposed in the literature. Wang et al[23] proposed the use of relation graphs to identify false data from injection attacks. Arvani and Rao[24] proposed a signal-based intrusion detection model to identify anomalies in the information reported by the sensors of cyber-physical systems. Do et al[19] presented the use of statistical detection methods, to identify data manipulation over stochastic processes. The use of watermark-based techniques are proposed by Mo et al[13] and revisited by Rubio-Hernan et al.[21,22] Next, we reexamine the watermark-based techniques presented in the works of Mo et al[13] and Rubio-Hernan et al[21,22] and show how to combine such techniques together with control strategy policies. We show that the resulting combination succeeds at handling the most powerful types of cyber-physical adversaries reported in the literature, ie, cyber-physical adversaries holding complete knowledge about the system dynamics.

3.2 | Control strategies

Control theory is a well-known topic, where the evolution of the technology has been the main motivation to create new control policies to manage these systems, keeping the control features. Among these new technologies, we can mention the networked control systems, where the loop between the different components of the system is closed through the network. The evolution of networked control systems to deal with the industry 4.0 is defined as cyber-physical industrial system. These new defined systems need the collaboration among a very wide number of disciplines to solve the challenges in term of autonomy, reliability, usability, functionality, and cyber security.[25]

Figure 1 shows a simple cyber-physical industrial system diagram, where f_s and f_c are the frequencies used by the sensors and controller, respectively, to process data. f_{ca} is the sampling frequency used in the communication channels between the controller and the actuators. f_{sc} is the sampling frequency used in the communication channels between the sensors and the controller.

Among control strategies used in cyber-physical industrial systems, we focus on strategies depending on the transmission policy necessary to secure the system. We assume policies not only to identify control faults but also intentional attacks carried out by external entities. Such policies can be classified as either sampled-data control policies or event-triggered control policies.

In the case of sampled-data control policies, the system can be driven using either monofrequency sampling (ie, same sampling frequency for all the channels) or multifrequency sampling (ie, different sampling frequencies depending on the channel, such as sensor-to-controller channel, or controller-to-actuator channel[7]). Figure 1, case 1, represents a monofrequency sampling scenario. Figure 1, case 2, represents a multifrequency sampling scenario.

Security in cyber-physical systems shall include the management of control properties and policies through the network, to avoid that an external entity, eg, an adversary, succeeds at harming the system. In the case of event-triggered control policies, we can differentiate between continuous event-triggered control[26] policies, where all the data received by the sensors are sent to the

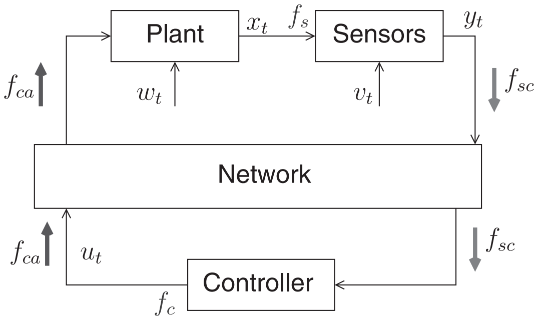


FIGURE 1 Cyber-physical system diagram. *Case 1*: $f_{ca} = f_{sc}$. *Case 2*: $f_{ca} \neq f_{sc}$. *Case 3*: $f_{ca} = f_c, f_{sc} = f_s$, and $f_s = f_c$. *Case 4*: $f_{ca} = \frac{f_c}{p_a}, f_{sc} = \frac{f_s}{p_s}$, such that $p_a, p_s \in \mathbb{N}$. *Case 5*: f_{sc} and f_{ca} →stochastic events

controller; periodic event-triggered control[8] policies, where sensors send update data periodically; and stochastic event-triggered schedule[9] policies, where sensors send update data only when there are variations in data obtained. Figure 1, cases 3, 4, and 5, represent each of these different types of policies.

3.3 | Watermark-based attack detection

We address in this section the watermark-based detector proposed by Mo et al.[13] The goal of such a detector is to identify cyber-physical replay attacks against networked control industrial systems. To analyze the watermark-based detector, Mo et al used an industrial control system modeled mathematically as a discrete linear time-invariant system. This mathematical model is used to describe the dynamic behavior of the system. The system can be represented as follows:

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{1}$$

$$y_t = Cx_t + v_t, \tag{2}$$

where $x_t \in \mathbb{R}^n$ is the state's vector, $u_t \in \mathbb{R}^p$ is the control signal, $y_t \in \mathbb{R}^m$ is the system output, and $w_t \in \mathbb{R}^n$ and v_t are the *process noise* and the measurement noise, respectively. The noises are assumed to be a zero mean Gaussian white noise with covariance Q, ie, $w_t \sim N(0, Q)$ and R, ie, $v_t \sim N(0, R)$, respectively. Moreover, $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times p}$, and $C \in \mathbb{R}^{m \times n}$ are the *state* matrix, the *input* matrix, and the *output* matrix, respectively. Let us now define the well-known *linear quadratic Gaussian* (LQG) approach used as a control technique in the work of Mo et al.[13] This technique has 2 independent components.

1. a *Kalman filter* producing an optimal state estimation $\hat{x}_t$ of the state x:

$$\begin{aligned} \hat{x}_{t|t-1} &= A\hat{x}_{t-1} + Bu_{t-1} \\ \hat{x}_t &= \hat{x}_{t|t-1} + K_t(y_t - C\hat{x}_{t|t-1}), \end{aligned} \tag{3}$$

   where K_t denotes the Kalman gain, and $\hat{x}_{t|t-1}$ is the *a priori* system state estimation.

2. a *linear quadratic regulator* providing the control law u_t.

$$u_t = L\hat{x}_t, \tag{4}$$

   where L denotes the feedback gain of a linear-quadratic regulator.

After describing the model of the plant, hereinafter, we present the detection scheme proposed in the work of Mo et al[13] against replay attacks.

The idea is to superpose a watermark signal $\Delta u_t \in \mathbb{R}^p$ to the optimal control law $u_t^\diamond$. The new control input u_t is given by

$$u_t = u_t^\diamond + \Delta u_t. \tag{5}$$

Note that the watermark signal is independent from the process noise w_t and the output noise v_t. To detect the adversaries, the watermark-based detector employs a well-known χ^2 detector.[27] Figure 2 shows the overall control system equipped with the attack detector proposed in the work of Mo et al.[13] The *alarm signal* g_t generated by the detector is defined as follows:

$$g_t = \sum_{i=t-w+1}^{t} (r_i)^T \mathcal{P}^{-1}(r_i), \tag{6}$$

where w is the size of the detection window, $\mathcal{P}$ is the covariance of input signals from the sensors, and $r_t = y_t - C\hat{x}_{t|t-1}$ is the residue generated from the estimator at each tth time step.

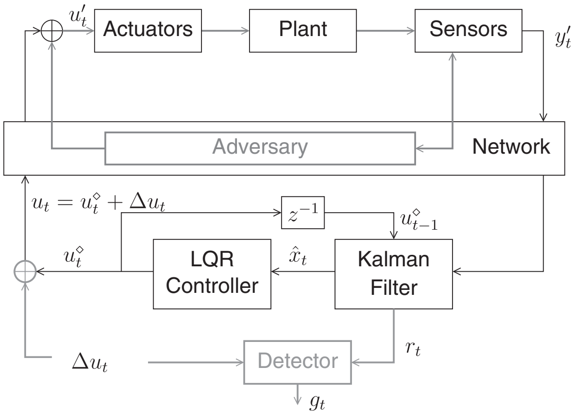


FIGURE 2 Watermark protection in cyber-physical systems.[13] LQR, linear quadratic regulator

To verify if the system is under attack, g_t is compared with threshold γ. If g_t is equal or greater than the threshold, $g_t \geq \gamma$, then the detector generates an alarm. The aforementioned threshold is chosen with respect to the system properties, eg, during the system identification phase, to keep a proper ratio of false positives and detector efficiency.

3.4 | Watermark-based attack detection against cyber-physical adversaries

The stationary challenge-response watermark-based detector proposed in the work of Mo et al[13] has been analyzed and revisited in the work of Rubio-Hernan et al.[21,22] In the work of Rubio-Hernan,[21] a new adversary, named nonparametric cyber-physical adversary is defined to show some weaknesses of using a stationary watermark in the detector. The adversary is able to use nonparametric identification techniques to identify the behavior of the system from the input and output data of the system, ie, from data sent to the actuator and data received from the sensors. This adversary bases the attack on the stationarity of the data. These stationary data allows the adversary to identify the correlation between the different events executed in the system. To handle such a vulnerability, a new watermark-based detector is proposed in the work of Rubio-Hernan et al.[21] The new approach uses different distributions to move toward an adaptive, and nonstationary, watermark-based detector with 2 main configurable parameters: the number of distributions N and the switching frequency $f = 1/T$. The new watermark is defined as follows:

$$\Delta u_t = \Delta u_t^{(s(t,T))}, \tag{7}$$

where $s : \mathbb{N} \times \mathbb{R} \rightarrow \mathcal{I}$, $\mathcal{I} = \{0, 1, \ldots, N-1\}$, is a static function that maps the time sample t and the switching period T to an element of the index set $\mathcal{I}$, defined as follows:

$$s(t,T) = \left\lfloor \frac{1}{T} \quad \text{mod} \quad (t, NT) \right\rfloor, \tag{8}$$

where mod(x, y) is the modulo operator, and $\lfloor \cdot \rfloor$ is the floor function.

The new watermark-based detector has been analyzed in the work of Rubio-Hernan et al[22] against a more powerful adversary. The new adversary, as in the case of the nonparametric adversary, uses parametric identification techniques to obtain a precise version of the system model. Unlike the nonparametric adversary, the techniques used by the new adversary aim at computing some precise system parameters. This allows to obtain the behavior of the system regardless of using nonstationary watermarks. To do so, the new adversary needs to guess 2 main configuration parameters: the precise number of samples to identify the system and the order of the system, ie, the dimension of the correlation matrix between the input and the output. This new adversary is able to evade detection if those 2 aforementioned parameters are properly chosen. In the following section, we present a new strategy to handle the adversary presented in the work of Rubio-Hernan et al.[22]

3.4.1 | Performance loss

The quadratic cost J of a LQG controller, with respect to the state x_t and the control input u_t, can be obtained using the following expression:

$$J = \lim_{n\to\infty} E\left[\frac{1}{n}\sum_{i=0}^{n-1}\left(x_i^T \Gamma x_i + u_i^T \Omega u_i\right)\right], \tag{9}$$

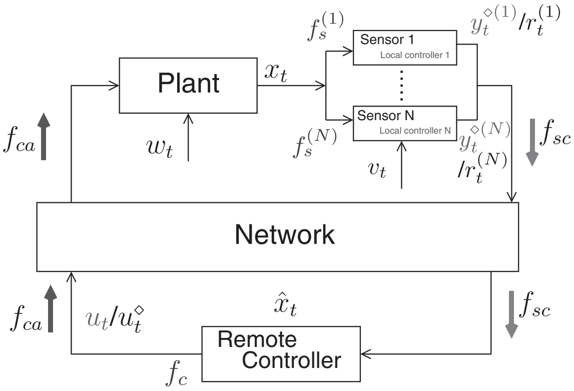


FIGURE 3 Cyber-physical system diagram with the new control security strategy

where Γ and Ω represent positive definite cost matrices.[28] The strategy proposed in this paper uses 2 detectors previously proposed in the works of Mo and Sinopoli[16] and Rubio-Hernan et al.[21] Both detectors use the theoretical methodology proposed by Mo et al,[13] to compute optimal detection, in terms of performance loss. Our approach does not increase the performance loss with respect to previous work.

4 | PIETC-WD WATERMARK-BASED DETECTION STRATEGY

In the previous section, we have seen that the watermark-based schemes are able to handle attacks carried out by adversaries with limited knowledge about the system dynamics, for instance, the ones defined as either cyber adversaries or nonparametric cyber-physical adversaries (cf Section 3.4). Nevertheless, it fails at detecting those adversaries with enough knowledge about the system dynamics, defined as parametric cyber-physical adversaries (cf Section 3.4). In this section, we present a new detector scheme, hereinafter denoted as periodic and intermittent event-triggered control watermark detector (PIETC-WD). This new detector aims at detecting cyber adversaries as replay attack and the cyber-physical adversaries presented in previous section.

Our scheme consists of a local controller located in each sensor and a remote controller creating a distributed controller (cf Figure 3). The cooperation between the local and the remote controller allows us to create an intrusion detection policy to capture integrity attacks. The local controllers manage the dynamics of the plant, and the remote controller manages the closed-loop system to ensure the system against integrity attacks. Notice that our new scheme requires an additional controller, together with the sensors, which must have enough computation power to process data estimations, eg, to predict errors between environmental and estimated data. The actuators do not require additional computational power. Nevertheless, during the time between 2 consecutive events, they must keep the last data received from the remote controller.

To perform with our scheme, it is necessary to define communication policies among the sensors, the actuators, and the remote controller. We define 2 communication policies for ensuring the system: (1) *periodic communication policy*, which the communication from the sensors to the remote controller is periodical, with a $T_{sc} = 1/f_{sc}$ period, and from the remote controller to the actuators, with a $T_{ca} = 1/f_{ca}$ period; and (2) *intermittent communication policy*, which allows for sending data from the sensors to the remote controller if the local controller produces an alarm. Notice that T_{sc} cannot be equal to T_{ca} to avoid that an intermittent communication takes place while the periodic communication is being sent.

Definition 1. Periodic and intermittent event-triggered control watermark detector is a detector strategy with distributed control tasks. On one hand, the sensors control the system periodically, using their local controllers and a local watermark-based detector.[13] On the other hand, the remote controller uses the estimation error received from each sensor to periodically generate the control inputs. The remote controller also controls the closed-loop communication with an intermittent watermark.

The PIETC-WD strategy uses essential properties from cyber-physical systems, to detect the attacks. The 3 main properties used in our work are as follows: (1) stability of closed-loop systems; (2) controllability of the system, to get full reachability; and (3) observability, to correlate events from the system. The periodic communication policy of PIETC-WD uses these 3 properties to verify if the system works properly, with regard to the information (eg, measurements) provided by the sensors. The intermittent communication policy verifies the stability of the closed loop from the controller standpoint, to avoid manipulation from external agents.

The communication algorithms, used to perform the PIETC-WD strategy, are shown in Algorithms 1 and 2. Algorithm 1 shows the local controller implementation, placed in the sensors, whose input is the data obtained from the physical system, y_t^i, and its output is the residue of the local controllers, r_t^i (with a challenge response watermark), if the alarm is disabled; or the value obtained by the sensors, y_t^i, if the alarm is enabled. Algorithm 2 shows the remote controller implementation, whose input is the data sent by the sensors, $data_{sc}$; and its output is the control inputs sent by the controller, $data_{ca}$, and the alarm value, $alarm_c$. More information about the controllers and the communication policies is provided in the sequel. The following notation will be used hereinafter:

$$\begin{aligned} \mathsf{U} &= \{u_0, u_{T_{ca}}, \dots, u_{\kappa T_{ca}}\} && \text{control input variables} \\ \mathsf{Y} &= \{y_0, y_{T_{sc}}, \dots, y_{\eta T_{sc}}\} && \text{observed measurements,} \end{aligned}$$

where $\alpha T_{ca} = T_{sc}$ with α, η and $\kappa \in \mathbb{N}$.

Algorithm 1 Communication policies in the sensors.

```
procedure SENSOR ALGORITHM
    $T_{sc} \leftarrow$ sensors/remote controller cycle
    $w_s \leftarrow$ sensor alarm window
top_s:
    $r_t^* \leftarrow y_t^i - \hat{y}_t^i$.
    $g_t^i \leftarrow \chi^2(r_t^*)$.
    if $g_t^i >$ Threshold then
        $alarm\_s \leftarrow alarm\_s + 1$.
    if $alarm\_s == w_s$ then
        $data_{sc} \leftarrow y_t$.
        $alarm\_s \leftarrow 0$.
    else
        if $mod(t, T_{sc}) == 0$ then
            $r_t^{(i)} \leftarrow y_t^{(i)} - C_i \hat{x}_{t|t-1}^{(i)}$.
            $data_{sc} \leftarrow r_t$.
    $t \leftarrow t + 1$.
    $\hat{y}_t^i = Local\_control\_feedback(r_{t-1}^*)$.
    goto top_s.
```

4.1 | Local controller design

The local controller is located in the sensors and uses a watermark to verify that the dynamics of the system is correct. Each sensor has a local controller with an LQG approach (cf Section 3.3). We denote the local controller in each sensor by $i \in \{0, 1, \dots, N-1\}$, where N is the number of sensors in the system. This controller adds a watermark to the sensor measurement before sending the residue to the remote controller

$$y_t^{(i)} = y_t^{\diamond(i)} + \Delta y_t^{(i)} \tag{10}$$

$$r_t^{(i)} = y_t^{(i)} - C_i \hat{x}_{t|t-1}^{(i)}, \tag{11}$$

where $y_t^{\diamond(i)}$ is the sensor measurement, $\Delta y_t^{(i)}$ is the watermark added by the local controllers, and $r_t^{(i)}$ is the residue sent to the remote controller to compute the control input $u_t = [u^{(0)}, \dots, u^{(N-1)}]$. Notice that the new sensor measurement $y_t^{(i)}$ is computed after verifying that $y_t^{\diamond(i)}$ is the correct sensor measurement.

4.1.1 | Remote controller design

The remote controller receives periodically the residue of each sensor, $r_t^{(i)}$, and computes these residues using the *LQG* approach (cf Section 3.3) to obtain the state estimation

$$\hat{x}_t = \hat{x}_{t|t-\varrho} + K_t(r_t) \tag{12}$$

and

$$r_t = r_\tau \qquad \forall t \in [\tau, \tau + (\alpha - 1)T_{ca}]\,, \tag{13}$$

where r_t is a vector generated by all the residues of the sensors and r_τ with $\tau \in [t \,:\, mod(t, T_{sc}) = 0]$ is the periodical residues' vector. In Equation 12, $\hat{x}_{t|t-\varrho}$ is defined as

$$\hat{x}_{t|t-\varrho} = A^{\varrho}\hat{x}_{t|t-\varrho} + \sum_{j=1}^{\varrho} A^{j-1}Bu_{\varrho-j+1}, \tag{14}$$

where $\varrho \in \{1, \ldots, \alpha\}$.

Algorithm 2 Communication policies in the controller.

```
procedure CONTROLLER ALGORITHM
    $T_{sc}$ ← sensors/remote controller cycle
    $T_{ca}$ ← remote controller/actuators cycle
    $data_{ca}$ ← data sent by the remote controller
    $data_{sc}$ ← data sent by the sensor
    $w_c$ ← controller alarm window
    $r_{data}$ ← reception data
top_c:
    if $r_{data} \neq True$ & $mod(t, T_{sc}) == 0$ then
        alarm_c ← alarm_c + 1.
        goto send_c.
    else
        if $r_{data} == True$ then
            if $mod(t, T_{sc}) == 0$ then
                $g_t^i \leftarrow \chi^2(r_t^{(i)})$.
                if $g_t^i > Threshold$ then
                    alarm_c ← alarm_c + 1.
                goto send_c.
            else
                if $\Delta u == 0$ then
                    alarm_c ← alarm_c + 1.
                else
                    $r_t^{(i)} \leftarrow data_{sc}^{(i)} - C_i\hat{x}_{t|t-1}^{(i)}$.
                    $g_t^i \leftarrow \chi^2(r_t^{(i)})$.
                    if $g_t^i > Threshold$ then
                        alarm_c ← alarm_c + 1.
        goto send_c.
send_c:
    if $alarm\_c == w_c$ then
        attack = True.
        alarm_c ← 0.
    if $mod(t, T_{ca}) == 0$ then
        $data_{ca} \leftarrow u_t$.
    $t \leftarrow t + 1$.
    goto top_c.
```

Equation 12 works properly if the data used by each sensor are independent among them or the system has only one sensor. Otherwise, if the data received by sensors are correlated, it is necessary to add a rectification for each sensor's residue to consider the cross-correlation among the watermarks added by local controllers. We can compute this rectification for $t \in [T_0, T_0 + T - 1]$ as follows:

$$r_t^{\text{rect}}(i) = r_t(i) - C_i\Delta\hat{x}_{t|t-T}^{i}\,, \tag{15}$$

where $r_t^{\text{rect}}(i)$ is the rectified value respect to different sensors. In addition, $\Delta\hat{x}^i_{t|t-T}$ is the rectification.

$$\Delta\hat{x}^i_{t|t-T} = \sum_{j=0}^{t-T_0-1} \left((A_i + B_iL_i)^{j+1} K_i(\epsilon_r) \right), \tag{16}$$

where $\epsilon_r = r_{t-1-j} - r^{(i)}_{t-1-j}$.

We can define the control inputs vector, u_t, as follows:

$$\begin{aligned} u_t &= L\left(\hat{x}_{t|t-\varrho} + K_t r_t^{\text{rect}}\right) \\ &= L\left(\hat{x}_{t|t-\varrho} + K_t(r_t^* + \Delta y_t)\right), \end{aligned} \tag{17}$$

where r_t^* is the vector of the residues before adding the watermark and their respective rectification, and Δy_t is the vector generated by all the sensors' watermarks.

The watermark used intermittently by the remote controller is added to the control inputs. The controller adds a watermark with probability β. Denoting $\lambda_t = 1$ or 0 as indication function whether the watermark is added or not, we assume that $\lambda' s$ are independent and identically distributed (iid). Bernoulli random variables with $E[\lambda_t] = \beta$.

The intermittence of the watermark communication allows us to define the watermark behavior as a nonstationary distribution. This watermark, Δu_t (cf Equation 5), permits us to detect if the closed loop is being manipulated. It is worth noting that Δu_t is a stochastic signal with the same variance as Δy_t.

4.2 | Periodic communication policy

The periodic communication policy is managed by the sensors. The sensors add the watermark in the measurements received by the plant and send residue r_t to the remote controller. The remote controller uses these residues to generate the control inputs sent to the actuators. The actions of these actuators produce change in the state of the plant that is captured by the sensors. If the real states differ from the state estimated by the sensors, then the sensors will switch from periodic communication policy to intermittent communication policy (cf Section 4.3).

In order to validate the proposal, let us assume that an attack is started at time T_0, and we compute residue $r_t^{(i)}$ for $t \in [T_0, T_0 + T - 1]$

$$r_t^{(i)} = y_t^{\prime(i)} - C_i\hat{x}^{(i)}_{t|t-T}, \tag{18}$$

where $y_t^{\prime(i)}$ is the sensor measurement sent to the controller by the adversary. Moreover, it is easy to show that the following holds:

$$\hat{x}^{(i)}_{t|t-T} = \hat{x}^{\prime(i)}_{t|t-T} + \mathcal{A}_i^{t-T_0}\left(\hat{x}^{(i)}_{T_0|T_0-1} - \hat{x}^{\prime(i)}_{T_0|T_0-1}\right) + \sum_{j=0}^{t-T_0-1}\left(\mathcal{A}^j(A_i + B_iL_i)K_i(\epsilon_{\Delta y})\right)$$

where $\epsilon_{\Delta y} = (\Delta y^{(i)}_{t-1-j} - \Delta y^{\prime(i)}_{t-1-j})$, $\hat{x}^{\prime(i)}$ is the local estimated state for each sensor when the system is under attack, and $\mathcal{A}_i = (A_i + B_iL_i)(I_i - K_iC_i)$ is a stable matrix.[13] Substitution of 19 in 18 yields

$$r_t^{(i)} = \underbrace{y_t^{\prime(i)} - C_i\hat{x}^{\prime(i)}_{t|t-T}}_{\text{First term}} - \underbrace{C_i\mathcal{A}_i^{t-T_0}\left(\hat{x}^{(i)}_{T_0|T_0-1} - \hat{x}^{\prime(i)}_{T_0|T_0-1}\right)}_{\text{Second term}} - \underbrace{C_i\sum_{j=0}^{t-T_0-1}\left(\mathcal{A}^j(A_i + B_iL_i)K_i(\epsilon_{\Delta y})\right)}_{\text{Third term}}.$$

Let us consider separately the 3 terms in the equation written above: the first term follows the same distribution of $(y_t - C_i\hat{x}^{(i)}_{t|t-1})$; since $\mathcal{A}_i$ is asymptotically stable, ie, all its eigenvalues are inside the open unit disk of the complex plane, the second term converges exponentially to zero. In fact, the entries of $\mathcal{A}_i^{t-T_0}$ converge exponentially fast to zero. The third term, under attack, is not equal to zero, since $\Delta y_t^{(i)} \neq \Delta y_t^{\prime(i)}$, and the adversary is detected; for a cyber adversary viewpoint, the measurements of the sensors change all the time and replay measurements are not accepted; likewise, a cyber-physical adversary is not able to obtain the system model using the methodology proposed in Section 3.4. The parametric cyber-physical adversary model, H_{at}, using the autoregressive with exogenous input identification approach,[29] is computed as follows:

$$H_{at} = f_{at}(u_t, r_t, y_t, v_t), \tag{19}$$

where f_{at} is a linear function that map the input u_t, as a function of t to the output (residue r_t, or measurement y_t) using a process noise, v_t, to measure a stochastic error.

Assuming that the real model is, $H = f(u_t, y_t, v_t)$, a linear function. Furthermore, the model shown by the system is $H = f_s(r_t, y_t, v_t, u_t, \Delta u_t, \Delta y_t)$, a non linear function, because we add an Δy_t in the measurements of the sensors periodically and an Δu_t in control input, as stochastic event, avoiding to follow a linear function to describe the system. For this reason, a powerful cyber-physical adversary is detected.

4.3 | Intermittent communication policy

The aforementioned periodic communication policy is managed by the sensors. The sensors produce an alarm if $g_t \geq \gamma$. When a sensor produces an alarm, this information is sent immediately to the remote controller. The affected sensor sends the real measurement to the remote controller to perform a second verification. This alarm is activated in a sensor if the control input has been manipulated by an external entity, a problem occurs in the system or the remote controller adds the watermark in the control input.

As we have defined before, a remote controller can generate an intermittent communication to verify that the closed loop is not broken. For a SISO (single-input, single-output) system or SIMO (single-input, single-output) system, the remote controller follows the next steps.

- The controller uses a low-pass filter after computing the control input, u_t, to remove the variations, and generate an optimal control input, u_t^*.

$$u_t^* = L\left(\hat{x}_{t|t-\varrho} + K_t r_t^*\right) \tag{20}$$

- The remote controller adds the owns watermark.

$$u_t = L\left(\hat{x}_{t|t-\varrho} + K_t r_t^*\right) + \Delta u_t \tag{21}$$

The local controllers located in the sensors receive the modification. They generate an alarm and send the measurement of the sensors to the remote controller. Then, the remote controller verifies if the measurement of the sensors is correct using the χ^2 detector. If the detector value is under the threshold, the system works correctly. Otherwise, if this value is over the threshold, the remote controller generates an alarm. In both cases, the next control input is computed as follows:

$$u_{t+T_{ca}} = L\left(\hat{x}_{t|t-\varrho} + K_t r_t\right), \tag{22}$$

where residue r_t follows Equation 13.

For a MISO (multiple-input, single-output) system or a MIMO (multiple-input, multiple-output) system the remote controller follows the next steps:

- The controller computes the control input:

$$\begin{aligned} u_t &= L(\hat{x}_{t|t-\varrho} + K_t r_t) \\ &= \left[u^{(1)}, \ldots, u^{(i)}, \ldots, u^{(n)}\right], \end{aligned} \tag{23}$$

  where n is the number of control inputs.
- Then the new control input vector is generated as follows:

$$u_t \begin{cases} u_t & = & \left[u^{(1)}, \ldots, u^{(i)}, \ldots, u^{(n)}\right] & \forall i \neq j \\ u_t^{(i)} & = & u_t^{*(i)} + \Delta u_t^{(i)} & i = j \end{cases},$$

  where $u_t^{*(j)}$ is the control input of the actuator, j, filtered to remove the variations. After removing the variation, the remote controller adds its watermark. It is worth noting that j is the actuator chosen by the controller to send the watermark.

As we have shown before, if the detector value, generated from the real data sent by the sensor, is under the threshold, the next control input is computed as follows:

$$u_{t+T_{ca}} = L(\hat{x}_{t|t-\varrho} + K_t r_t). \tag{24}$$

Otherwise, the remote controller generates an alarm.

Briefly, when the remote controller receives a measurement from a sensor, if a watermark Δu has not been sent, then the remote controller creates an intrusion alarm. Otherwise, if a watermark has been added to the control input, the controller verifies if this alarm is produced by the watermark. If the residue generated between the real measurements of the sensors and the estimation is under the threshold, the remote controller sends the control input generated before adding the watermark. However, if the residue is over the threshold, it means that an external entity is into the closed loop, and an alarm is activated.

In order to validate our claims, let us assume the following attack in the communication channel between the sensor and the controller after the controller sends a control input with a watermark. It is started at time T_0, and the remote controller includes the remote watermark, Δu_t at time $T_1 \in [T_0, T_0 + T - 1]$. We compute the residues r_t for $t \in [T_0, T_0 + T - 1]$:

$$r_t \begin{cases} r_t = & \left[r^{(1)}, \ldots, r^{(i)}, \ldots, r^{(n)}\right] \qquad t \in [T_0, T_1] \\ r_t = & y'_t - C\hat{x}_{t|t-T} \qquad t \in [T_1, T_0 + T - 1] \end{cases}. \tag{25}$$

On one hand, it is worth noting that the attack is not detected between T_0 and T_1, since the adversary is able to replay or insert a correct residue without being detected. On the other hand, for $t \in [T_1, T_0 + T - 1]$, it is easy to show that the following holds:

$$\hat{x}_{t|t-T} = \hat{x}'_{t|t-T} + \mathcal{A}^{t-T_1}\left(\hat{x}_{T_1|T_1-1} - \hat{x}'_{T_1|T_1-1}\right) + \sum_{j=0}^{t-T_1-1}\left(\mathcal{A}^j B\left(\Delta u_{t-1-j} - \Delta u'_{t-1-j}\right)\right). \tag{26}$$

Substitution of (26) in (25) yields:

$$r_t = \underbrace{y'_t - C\hat{x}'_{t|t-T}}_{\text{First term}} - \underbrace{C\mathcal{A}^{t-T_1}\left(\hat{x}_{T_1|T_1-1} - \hat{x}'_{T_1|T_1-1}\right)}_{\text{Second term}} - \underbrace{C\sum_{j=0}^{t-T_1-1}\left(\mathcal{A}^j B\left(\Delta u_{t-1-j} - \Delta u'_{t-1-j}\right)\right)}_{\text{Third term}}.$$

The first term follows the same distribution of $(y_t - C\hat{x}_{t|t-1})$; the second term converges exponentially to zero. Since the third term is not equal to zero, $\Delta u_t \neq \Delta u'_t$, the adversary is detected; from the cyber adversary viewpoint, the measurements of the sensors change all the time and replay measurements are not accepted; likewise, the cyber-physical adversary is not able to obtain the system model.

4.4 | New parametric cyber-physical adversary

In this section, we present a new parametric cyber-physical adversary with the knowledge about the new detector strategy, to evaluate new detection strategy. This attacker has knowledge about the new communication policies and the existence of the local and the remote watermarks. Nevertheless, the new adversary does not know the watermark covariances, the controller's parameters used to obtain the correct error between data, and neither the moment when the remote controller forces an intermittent communication.

The new adversary could be able to detect the correlation model between the inputs and the outputs of the plant. This adversary can force the sensors' intermittent communication with malfunction control inputs, and mislead the controller with replay error data to obtain the model. Nevertheless, this adversary is not able to know when the communication is periodic or intermittent, since the attacker does not know when the remote control sends the watermark added to the control inputs, which generate the intermittent communication. The intermittent communication does not change the communication between the remote controller and the actuators, but produces an intermittent communication between the sensors and the remote controller, necessary to verify the closed loop.

Briefly, the new adversary is able to attack the integrity of the system. Nevertheless using the PIETC-WD strategy, the adversary is detected by the controllers of the sensors. The remote controller detects the attack when it verifies the behavior of the closed loop. The adversary cannot avoid the alarm in the sensors (local controller). Nevertheless, the attacker can cut off the communication between the sensors and the remote control misleading the remote controller with correct residues (eg, replay residues). Moreover, to avoid the alarm in the remote controller, the adversary can switch between sending the measurements of the sensors or the residues, but the adversary has a great probability to be detected. We validate the PIETC-WD strategy against the new parametric cyber-physical adversary in the next section.

5 | NUMERICAL VALIDATION

This section presents a numerical validation of the strategy. We have simulated a multiple-input–multiple-output system, which represents a cyber-physical system with 4 inputs and 4 outputs. To define mathematically this system using the model shown

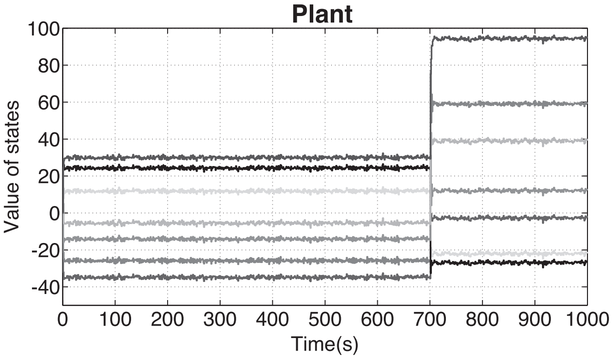


FIGURE 4 The dynamics of the states vector in the plant under a parametric attack. The attack starts at $t = 700s$

in Section 3.3, we use the following matrices*:

$$A = \begin{bmatrix} -0.199 & 0.145 & -0.01 & -0.119 & 0.062 & 0.134 & 0.001 \\ 0.0413 & -0.189 & -0.124 & 0.037 & 0.167 & 0.059 & 0 \\ 0.031 & -0.097 & -0.344 & -0.055 & -0.138 & 0.213 & -0.002 \\ -0.156 & -0.030 & 0.042 & -0.014 & 0.005 & 0.143 & 0 \\ -0.016 & 0.162 & -0.129 & 0.118 & -0.199 & 0.005 & 0.001 \\ 0.096 & 0.082 & 0.203 & 0.042 & -0.028 & -0.147 & 0 \\ -0.250 & 0.087 & -0.064 & -0.134 & -0.093 & 0.075 & 0.447 \end{bmatrix},$$

$$B = \begin{bmatrix} -1.739 & 0 & 0 & 0 \\ -0.214 & 0 & 0.702 & 0 \\ 1.437 & 0.001 & -1.6 & -0.711 \\ -1.107 & 0.002 & 0 & 1.548 \\ 0 & 0 & -0.952 & 0 \\ 1.197 & 0 & -1.01 & 0 \\ 0.953 & 0.765 & -1.44 & 0 \end{bmatrix},$$

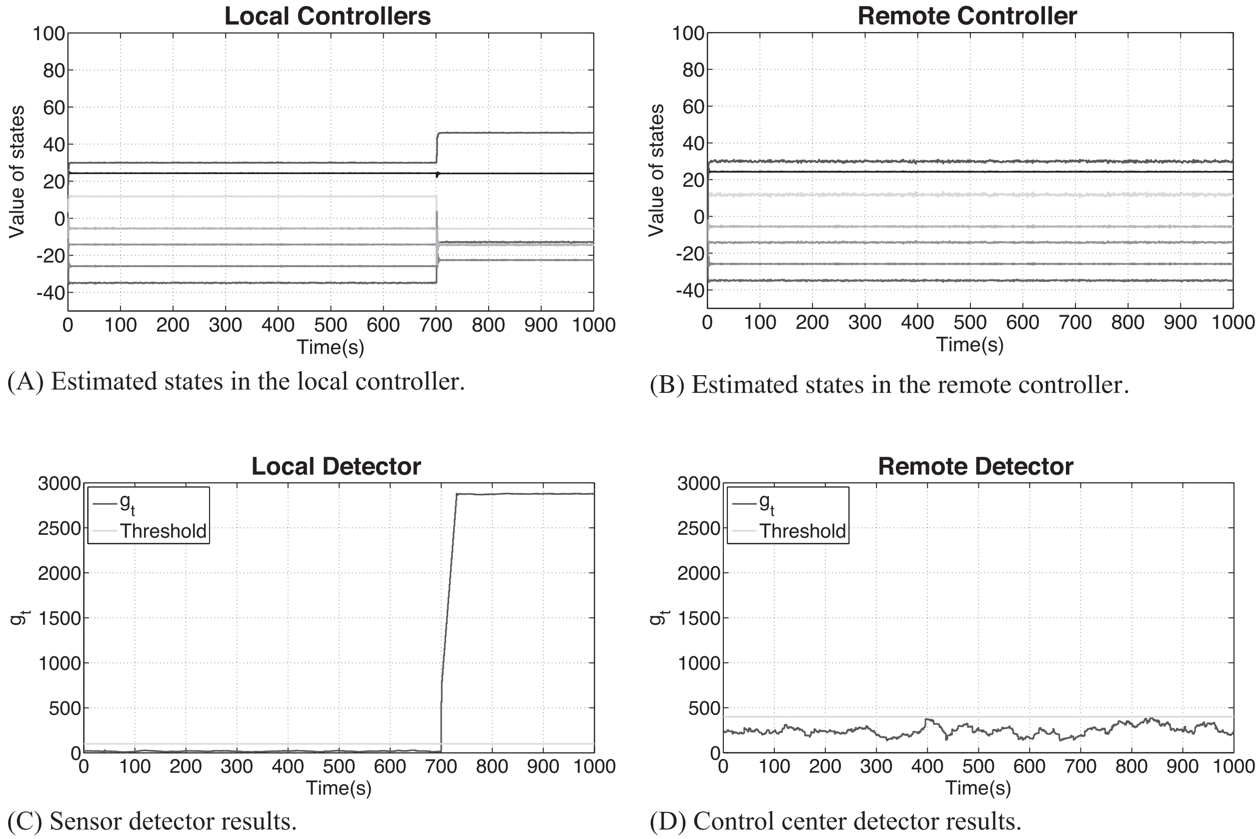
$$C = \begin{bmatrix} -2.261 & 1.112 & -0.09 & 0 & 0.457 & 0.829 & -1.137 \\ 0.091 & 0.259 & 0.876 & -0.273 & 2.236 & -1.484 & -1.091 \\ 0 & 0.689 & -0.074 & 0 & -0.111 & 0 & -0.310 \\ 0 & 0.524 & -0.195 & 0 & 0 & -0.138 & 0 \end{bmatrix}.$$

We present in Figure 4 the dynamic of the plant and how the adversary can modify this dynamic to disrupt the system. Figure 5 shows how the detectors present in the sensors are able to detect attacks using the periodical communication policy and a watermark managed by sensors. Figure 5A,B shows the dynamic of the states estimated by the local controllers placed in sensors and the remote controller placed in control center, respectively. Figure 5C,D shows the dynamic of the alarm in sensors (local detectors) and in the control center (remote detector). We show in these Figures an attack, which is detected by the local detectors, but whose alarm's dynamic is not transmitted to the control center. The sensors' alarms are stopped by a cyber-physical adversary that takes the control of the system and sends a dynamic alarm correlated with the data sent from the remote controller (control center) to the actuator. This action allows the adversary attack the system without be detected.

Figure 6 presents the control/security strategy using the periodical and intermittent communication policies. As shown in Figure 6, it is visible that using both policies allows the system to detect the attack. Figure 6C shows the local detector that is able to detect a disruption in the dynamic of the plant, thanks to the periodical communication policy. In Figure 6D, we can see that the remote detector (placed in the control center) is able to detect that the closed loop is broken, using the remote controller's watermark and the intermittent communication policy.

Our strategy handles integrity attacks at the physical level, by using the watermark sent by the local controller, ie, placed in the sensors (cf Figure 6A), as well as integrity attacks carried out through the network, via the closed-loop verification conducted by the remote controller, using stochastic metrics to trigger the watermark (cf Figure 6B). The states estimated by the local and remote controllers (cf Figure 6A,B) are handled and mitigated by the system, reporting the events to security operators (eg, in terms of alerts) and activating the appropriate corrective mechanisms.

*The covariance matrices are equal to $Q = 0.2I$ and $R = I$.



(A) Estimated states in the local controller.

(B) Estimated states in the remote controller.

(C) Sensor detector results.

(D) Control center detector results.

FIGURE 5 Numeric simulation results, using only periodical policy. Attacks start at $t = 700s$. A,B, The dynamics of the states vector estimated in the local and remote controller, respectively, under the same scenarios. C,D, The dynamics of the alarm signal g_t produced in the local and remote detector, respectively, under the same scenarios

6 | EXPERIMENTAL RESULTS

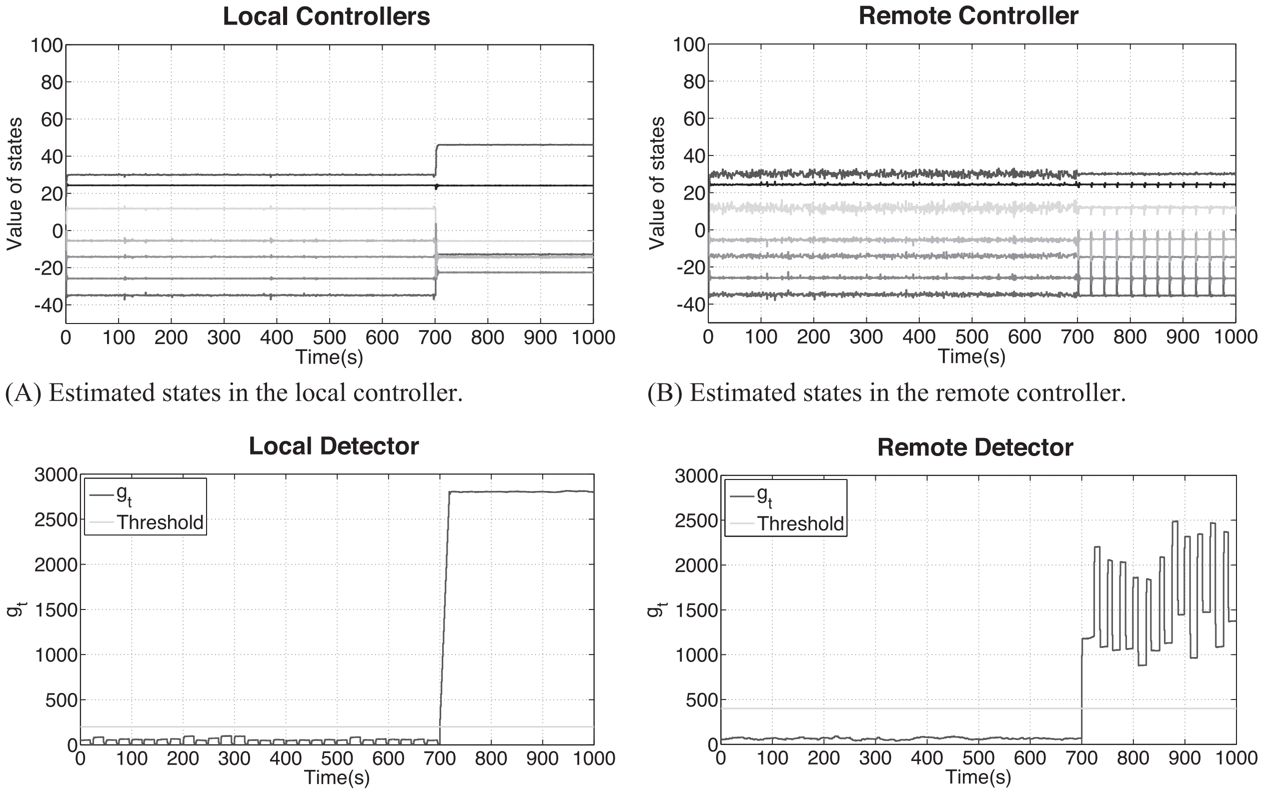
We present some experimental results obtained with a laboratory testbed. The testbed implements the detection strategy and the adversary models presented in this paper. The implementation is conducted using Modbus[30] and DNP3,[31] 2 well-known SCADA protocols used in cyber-physical industrial systems.

6.1 | SCADA Testbed

SCADA is a general term that encompasses well-defined types of field devices, such as: (1) master terminal units and human machine interfaces, located at the topmost layer and managing all communications; (2) remote terminal units (RTUs) and programmable logic controllers (PLCs), controlling and acquiring data from remote equipment and connecting with the master stations; and (3) sensors and actuators.

The design of our testbed includes all those aforementioned field devices: controllers, sensors, actuators, PLCs, and RTUs. The devices are distributed across several nodes in a shared network combining Modbus[30] and DNP3.[31] One or various field devices can be embedded into a single node. From a software standpoint, the controller never connects directly to the sensors. Instead, it is integrated in the testbed as a PLC node, with eventual connections to some other intermediary nodes. To run our experiments, we created a distributed network where the control center communicates with the physical devices through the RTUs. The communication between the control center and the RTUs uses the protocol DNP3 and the communication between the RTUs and the physical devices uses Modbus. The testbed can be expanded to handle additional SCADA protocols, as well as to connect complementary devices, such as additional PLCs and RTUs.

From a data transmission standpoint, the devices connected to the testbed use different sampling frequencies. The testbed is able to handle several PLCs. To avoid overloading 1 channel with all the possible registers of the PLCs, different ports are designated to isolate the communication between separated PLCs. DNP3 commands perform an integrity scan that gathers all the data from the PLCs in case several PLCs were being handled in the same channel. This avoids variables of the PLCs causing communication overheads.



(A) Estimated states in the local controller.

(B) Estimated states in the remote controller.

(C) Sensor detector results.

(D) Control center detector results.

FIGURE 6 Numeric simulation results, using periodical and intermittent policies. Attacks start at $t = 700s$. A,B, The dynamics of the states vector estimated in the local and remote controller, respectively, under the same scenarios. C,D, The dynamics of the alarm signal g_t produced in the local and remote detector, respectively, under the same scenarios

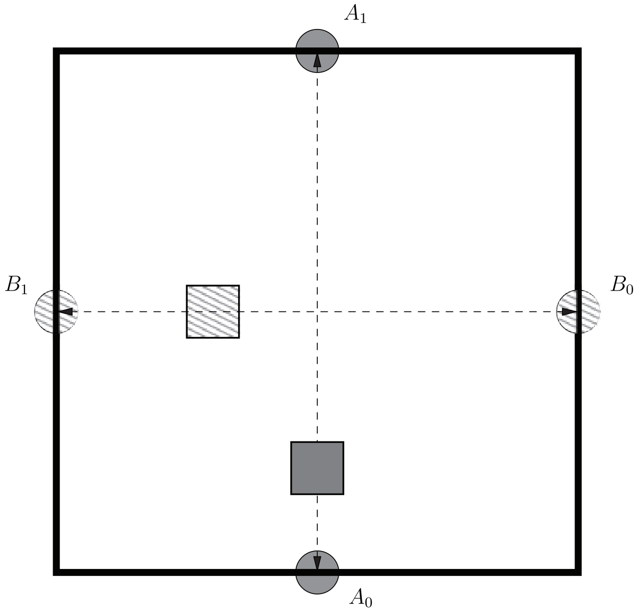
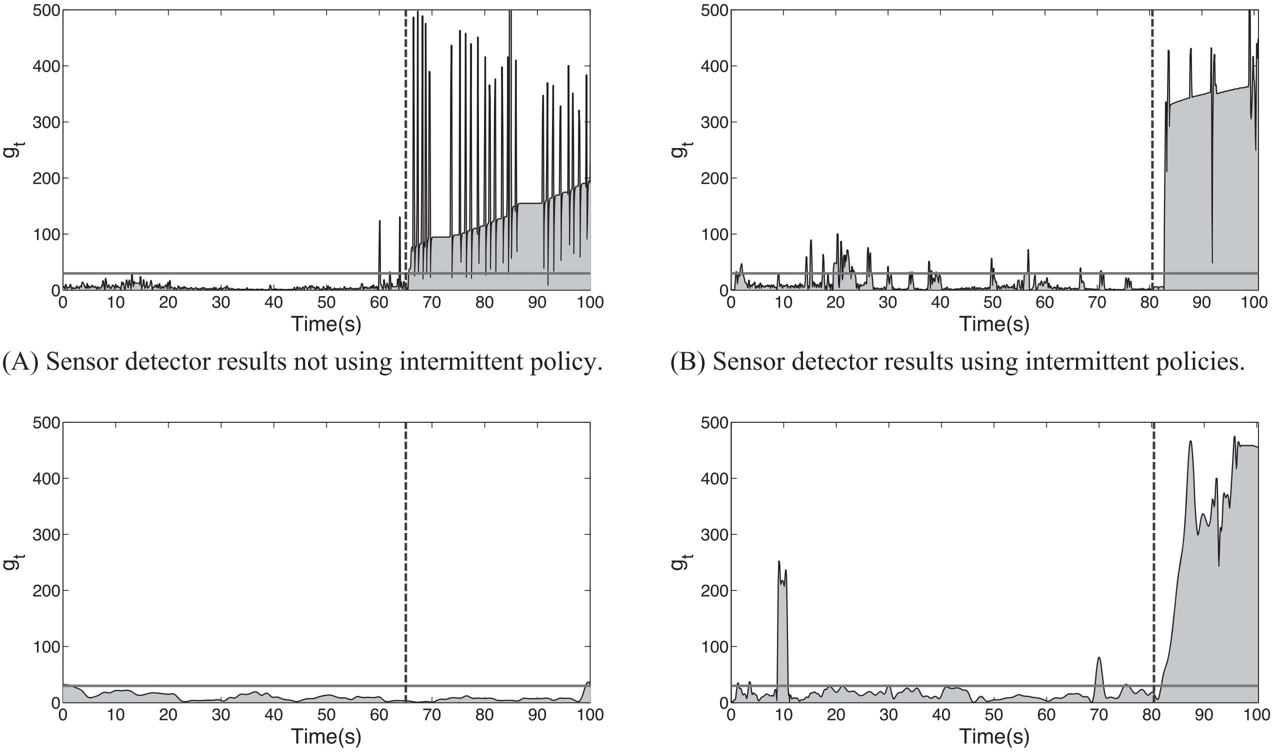


FIGURE 7 Cyber-physical industrial scenario implemented in our experimental testbed. The 2 mobile agents, represented by solid and pattern gray squares, move from and toward 2 spatial coordinates, represented by solid and pattern gray circles. Some live demonstration videos of this setup are available at http://j.mp/TSPScada



(A) Sensor detector results not using intermittent policy.

(B) Sensor detector results using intermittent policies.

(C) Control center detector results not using intermittent policy.

(D) Control center detector results using intermittent policy.

FIGURE 8 Experimental testbed results. The horizontal solid line represents the threshold. The vertical dotted line represents the moment when the attack starts. Peaks on the left side of the vertical dotted line represent false positives. A,B, Local detector values g_t (placed within the sensors) without and with intermittent policies (ie, remote watermarks) under a parametric attack. C,D, Remote detection values (placed at the control center) without and with intermittent policies under a parametric attack

A sample cyber-physical industrial scenario implemented in our testbed is shown in Figure 7. In the Figure, the solid and pattern gray squares represent 2 mobile agents that move from 2 spatial coordinates (represented by the the solid and pattern gray circles), and vice versa. Each mobile agent controls the dynamics of its movements. The 2 agents are coordinated by a SCADA controller, to avoid spatial collisions. Further information, as well as some live demonstration videos, is available at http://j.mp/TSPScada.

6.2 | Experimentation and results

We present the results of applying the detection strategy, assuming attackers under the parametric adversary model. The attacker tries to evade the detection techniques reported in this paper. In a first scenario, the system uses only a periodic communication policy using the watermark detector placed at the sensors (cf Figure 8A,C). When the attacker successfully infers the system parameters, the attack is not reported by the remote controller. Nevertheless, local alarms placed at the sensors report the attack. In a second scenario, the system uses also the intermittent policy. In such a case, the attack is detected (cf Figure 8B,D) by the remote controller. In Figure 8B, we may observe some peaks (cf, $t = [2.1s; 14.4s; 15.3s; 17.6s; 20.4s; 21.0s; 22.3s; 26.2s; 37.8s; 56.8s; 66.8s]$), representing the local detector's reaction in the presence of a watermark sent by the remote controller — used to verify that the closed loop works properly, as defined in Section 4.3.

To compare the watermark-based detector mechanism and our detection strategy, several repetitions of the experiment were orchestrated. Using the results obtained from the testbed experiments, we report in Table 3 the percentage of false positives, false negatives, and detection ratio. We also show the average time that the detector placed at the control center, takes to launch the alerts. Notice that the workload in the sensors, due to adding the local controller in our detection strategy, does not significantly increase the performance loss of the system. The increase, in terms of memory and processing time, is shown in Table 2.

Regarding the results shown in Table 3, we can emphasize that: (1) a system that uses only the watermark-based detector mechanism against a parametric attacker has a lower detection ratio, of about 12%. That is possible since the attacker can evade

TABLE 2 Memory and processing time increase, due to the PIETC-WD strategy implementation

	Time	RAM Memory
Sensor's workload increase	17.13 ms	5.23%

Abbreviation: PIETC-WD, periodic and intermittent event-triggered control watermark detector.

TABLE 3 Detector performance results

	Using Only Watermark Detector Mechanism	Using PIETC-WD Strategy
Detection ratio	12.00%	65.00%
Average detection time	6.08 s	7.30 s
False negatives	88.60%	35.00%
False positives	1.35%	3.5%

Abbreviation: PIETC-WD, periodic and intermittent event-triggered control watermark detector.

the detection process if they succeed at properly identifying the system attributes. This result is expected, as suggested by the theoretical and simulation-based conclusions presented in the work of Rubio-Hernan[22]; and (2) a system that uses the strategy proposed in this paper has a higher detection ratio, of about 65.00%. In this scenario, the detection ratio increases, thanks to the distributed detector proposed in Section 4, confirming the theoretical and simulation-base result reported in the previous sections. Moreover, the false negative ratio significantly decreases, from 88.60% to 35.00%. In terms of false positives, both scenarios show similar data, but the PIETC-WD strategy generates about 2.0% more. The time between the beginning of the attack and the moment, when the attack is detected by the remote controller, also increases, since the detection watermark managed by the remote controller follows a stochastic law.

7 | CONCLUSION

We have addressed security issues in cyber-physical industrial systems. We have presented an adaptive strategy that detects adversaries that are able to acquire knowledge about the system dynamics, prior to starting their attacks to get control over the inputs and measurements of the system. In addition to industrial systems, any other cyber-physical scenario, modeled in terms of sensors, actuators and feedback control, can benefit from the approach. Computer networks complemented by physical systems, such as Internet of things scenarios, may also benefit from the approach. We have validated the construction using numerical validation and a laboratory testbed based on SCADA industrial protocols. We have shown that the strategy is able to detect the attacks with a high detection ratio.

ACKNOWLEDGEMENTS

The authors acknowledge support from the Cyber CNI Chair of the Institut Mines-Télécom (cf http://chaire-cyber-cni.fr/). The chair is supported by Airbus Defence and Space, Amossys, EDF, Orange, La Poste, Nokia, Société Générale and the Regional Council of Brittany. It has been acknowledged by the Center of excellence in Cybersecurity.

REFERENCES

1. Hespanha JP, Naghshtabrizi P, Xu Y. A survey of recent results in networked control systems. *Proc IEEE*. 2007;95(1):138-162. https://doi.org/10.1109/JPROC.2006.887288
2. Kim KD, Kumar PR. Cyber-physical systems: a perspective at the centennial. *Proc IEEE*. 2012;100:1287-1308. https://doi.org/10.1109/JPROC.2012.2189792
3. Wu G, Sun J, Chen J. A survey on the security of cyber-physical systems. *Control Theory Technol*. 2016;14(1):2-10. https://doi.org/10.1007/s11768-016-5123-9
4. Falliere N, Murchu LO, Chien E. W32 stuxnet dossier. *White paper, Symantec Corp, Security Response*. 2011;5:6.
5. Weyer S, Schmitt M, Ohmer M, Gorecky D. Towards industry 4.0 - Standardization as the crucial challenge for highly modular, multi-vendor production systems. *IFAC-PapersOnLine*. 2015;48(3):579-584. https://doi.org/10.1016/j.ifacol.2015.06.143

6. Lee J, Bagheri B, Kao HA. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf Lett*. 2015;3:18-23. https://doi.org/10.1016/j.mfglet.2014.12.001
7. Salt J, Casanova V, Cuenca A, Pizá R. Sistemas de control basados en red modelado y diseño de estructuras de control. *Revista Iberoamericana de Automática e Informática Industrial RIAI*. 2008;5(3):5-20. https://doi.org/10.1016/S1697-7912(08)70157-2
8. Heemels W, Donkers M, Teel AR. Periodic event-triggered control for linear systems. *IEEE Trans Autom Control*. 2013;58(4):847-861. https://doi.org/10.1109/TAC.2012.2220443
9. Han D, Mo Y, Wu J, Weerakkody S, Sinopoli B, Shi L. Stochastic event-triggered sensor schedule for remote state estimation. *IEEE Trans Autom Control*. 2015;60(10):2661-2675. https://doi.org/10.1109/TAC.2015.2406975
10. Corman D, Pillitteri V, Tousley S, Tehranipoor M, Lindqvist U. NITRD cyber-physical security panel. Paper presented at: 35th IEEE Symposium on Security and Privacy, IEEE SP 2014, May 18-21, 2014; San Jose, CA, USA.
11. Mo Y, Garone E, Casavola A, Sinopoli B. False data injection attacks against state estimation in wireless sensor networks. Paper presented at: 49th IEEE Conference on Decision and Control (CDC); 2010; Atlanta, GA, USA. https://doi.org/10.1109/CDC.2010.5718158
12. Smith RS. Covert misappropriation of networked control systems: presenting a feedback structure. *IEEE Control Syst*. 2015;35(1):82-92. https://doi.org/10.1109/MCS.2014.2364723
13. Mo Y, Weerakkody S, Sinopoli B. Physical authentication of control systems: designing watermarked control inputs to detect counterfeit sensor outputs. *IEEE Control Syst*. 2015;35(1):93-109. https://doi.org/10.1109/MCS.2014.2364724
14. Teixeira A, Shames I, Sandberg H, Johansson KH. A secure control framework for resource-limited adversaries. *Automatica*. 2015;51:135-148. https://doi.org/10.1016/j.automatica.2014.10.067
15. Pasqualetti F, Dorfler F, Bullo F. Cyber-physical security via geometric control: Distributed monitoring and malicious attacks. Paper presented at: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC); 2012; Maui, HI, USA. https://doi.org/10.1109/CDC.2012.6426257
16. Mo Y, Sinopoli B. Secure control against replay attacks. Paper presented at: Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing. IEEE; 2009; Monticello, IL, USA. https://doi.org/10.1109/ALLERTON.2009.5394956
17. Miao F, Pajic M, Pappas GJ. Stochastic game approach for replay attack detection. Paper presented at: Proceedings of the 52nd IEEE Conference on Decision and Control; 2013; Florence, Italy. https://doi.org/10.1109/CDC.2013.6760152
18. Zhu Q, Başar T. Dynamic policy-based IDS configuration. Paper presented at: Proceedings of the 48th IEEE Conference on Decision and Control CDC) held jointly with 2009 28th Chinese Control Conference; 2009; Shanghai, China. https://doi.org/10.1109/CDC.2009.5399894
19. Do VL, Fillatre L, Nikiforov I. A statistical method for detecting cyber/physical attacks on SCADA systems. Paper presented at: Proceedings of 2014 IEEE Conference on Control Applications (CCA); 2014; Juan Les Antibes, France. https://doi.org/10.1109/CCA.2014.6981373
20. Genge B, Kiss I, Haller P. A system dynamics approach for assessing the impact of cyber attacks on critical infrastructures. *Int J Critical Infrastruct Prot*. 2015;10:3-17. https://doi.org/10.1016/j.ijcip.2015.04.001
21. Rubio-Hernan J, De Cicco L, Garcia-Alfaro J. Revisiting a watermark-based detection scheme to handle cyber-physical attacks. Paper presented at: Proceedings of the 11th International Conference on Availability, Reliability and Security. IEEE; 2016; Salzburg, Austria.
22. Rubio-Hernan J, De Cicco L, Garcia-Alfaro J. Event-triggered watermarking control to handle cyber-physical integrity attacks. Paper presented at: Proceedings Secure IT Systems: 21st Nordic Conference, Nordsec 2016, Oulu, Finland, November 2-4, 2016; Springer International Publishing: Cham. https://doi.org/10.1007/978-3-319-47560-8_1
23. Wang Y, Xu Z, Zhang J, Xu L, Wang H, Gu G. SRID: State relation based intrusion detection for false data injection attacks in SCADA. In: Kutyłowski M, Vaidya J, eds. *Computer Security - Esorics 2014: 19th European Symposium on Research in Computer Security, Wroclaw, Poland, September 7-11, 2014. Proceedings, Part II*. Cham: Springer International Publishing; 2014:401-418. https://doi.org/10.1007/978-3-319-11212-1_23
24. Arvani A, Rao VS. Detection and protection against intrusions on smart grid systems. *Int J Cyber Secur Digit Forensics (IJCSDF)*. 2014;3(1):38-48.
25. Baheti R, Gill H. Cyber-physical systems. *Impact Control Technol*. 2011;12:161-166.
26. Heemels W, Donkers M. Model-based periodic event-triggered control for linear systems. *Automatica*. 2013;49(3):698-711. https://doi.org/10.1016/j.automatica.2012.11.025
27. Brumback B, Srinath M. A chi-square test for fault-detection in Kalman filters. *IEEE Trans Autom Control*. 1987;32(6):552-554. https://doi.org/10.1109/TAC.1987.1104658
28. Franklin GF, Powell JD, Workman ML. *Digital control of dynamic systems*. 3rd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.; 1998.
29. Natke HG. System identification: Torsten Söderström and Petre Stoica. *Automatica*. 1992;28(5):1069-1071.
30. Modbus Organization. Official Modbus Specifications. http://www.modbus.org/specs.php. 2016. Accessed: March 2017.
31. Curtis K. A DNP3 protocol primer. A basic technical overview of the protocol. http://www.dnp.org/AboutUs/DNP3%20Primer%20Rev%20A.pdf. 2005. Accessed: March 2017.