Non-stationary Watermark-based Attack Detection to Protect Cyber-Physical Control Systems

Jose Rubio-Hernan*†, Luca De Cicco ‡, Joaquin Garcia-Alfaro†
E-mails: jose.rubio_hernan@telecom-sudparis.eu,luca.decicco@poliba.it,
joaquin.garcia_alfaro@telecom-sudparis.eu
† Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France
‡ Politecnico di Bari, Dipartimento di Ingegneria Elettrica e dell'Informazione, Bari, Italy

Abstract

This chapter addresses security issues in cyber-physical industrial systems. Attacks against these systems shall be handled both in terms of safety and security. Networked control technologies imposed by industrial standards already cover the safety dimension. From a security standpoint, the literature has shown that using only cyber information to handle the security of cyber-physical systems is not sufficient, since physical malicious actions, that can threaten the correct performance of the systems, are ignored. For this reason, cyber-physical systems should be protected from threats to their cyber and physical layers. Some authors handle the attacks by using physical attestations of the underlying processes. For instance, the use of physical watermarking can complement the protection techniques at the cyber layer, in order to ensure the truthfulness of the process. These detectors work properly if the adversaries do not have enough knowledge to mislead cross-layer (e.g., cyber and physical) data. Nevertheless, adversaries able to acquire enough knowledge from both layers may evade detection.

The solutions listed in this chapter handle those aforementioned limitations. The chapter starts by showing shortcomings of classical stationary watermark-based fault detectors, extended to detect, in addition to failures, malicious actions. It is shown that classical stationary watermark-based detectors are unable to identify cyber-physical adversaries. Specifically, they may only detect adversaries that do not attempt to get additional knowledge about the system dynamics. An analysis about the performance of a specific stationary watermark-based fault detector is presented. A new threat model is assumed, in which adversaries may now infer system dynamics by correlating both cyber and physical data. The goal of such adversaries is to evade detection. Under this new threat model, adversaries can now evade detection with high probability. To handle the issue, an extended strategy is presented. The idea is to transform the classical (stationary) approach into a non-stationary watermark-based detector. The new design is shown to handle the extended threat model. It is also shown new ways to combine control and communication strategies, to boost the detection performance. The new solutions are validated using both numeric simulations and cyber-physical testbeds. Ideas for future work are also presented.

Keywords: Cyber-Physical System, Cyber-Physical Security; Watermarking; Control Theory.

1 Introduction

Nowadays, an ever increasing number of companies and industrial facilities require to access critical data from any location, ensuring the control of both data and processes. This need makes the combination of network security and industrial control security a key research topic. Fields involved in this research area are: (1) Information and Communications Technology (ICT), which encompasses the control of computer networks and communication; (2) traditional cybersecurity, focused on creating detection techniques and countermeasures against attacks in the cyber domain; (3) Industrial Control Systems (ICS), focused on designing controllers to make the physical processes behave in such a way that specific static and dynamic performance metrics are met; and (4) safety in industrial processes, focused on methodologies to avoid failures and accidents in the process.

We define the terminology used throughout this chapter: (i) the terms *cyber* and *physical* are used to refer to second upper layers in information and communication technologies and control systems respectively; (ii) *cyber-physical systems* are the new generation of systems that combine cyber and physical components using data in the digital and continuous domain [33]; (iii) *Networked Control Systems*, which are a subset of cyber-physical systems dedicated to industrial control systems; and (iv) *Control Systems*, defined as an interconnection of components that form a physical system (often referred to as *factories* or *physical environment*) designed to provide a given desired response under a control law. We focus specifically on *closed-loop control systems*, that are control system requiring feedback from sensors to continuously compute the control actions in order to reach the control goals.

The security of cyber-physical systems is attracting the attention of the research community [11], especially after the *StuxNet* malware [14] revealed the potential of security attacks against such systems.

Several authors have investigated the requirements for addressing emerging security issues when designing security mechanisms for cyber-physical systems. In [8], Cardenas *et al.* define the security issues in these systems by analyzing the problem separately,

first from an information security perspective, and then by examining specific control issues. In [9], Cardenas *et al.* describe for the first time the difference between the security of conventional enterprise networks, and the security of cyber-physical systems. Figure 1 shows how adversaries conducting a cyber-physical attack can be represented as a block diagram scheme, a classical representation employed in the control systems community. The symbol $\bigoplus$ in the figure represents a *summing junction*, i.e., a linear element that outputs the algebraic sum of a number of input signals. The figure represents a closed-loop control system and how the adversaries are able to disrupt the nominal behaviour of the system. Specifically, the adversaries modify the control input u_t (by inserting u'_t instead) to affect the state of the system and disrupt the normal operating conditions. Adversaries do not need the knowledge of the system's process model. However, access to all sensors (i.e., all components of the y_t vector) or communication protocols is required to perform an attack, i.e., to be able to insert the correct y_t vector in place of the y'_t vector generated due to the malicious u'_t data. This type of adversary is then undetectable with a detector that only checks for faulty measurements.

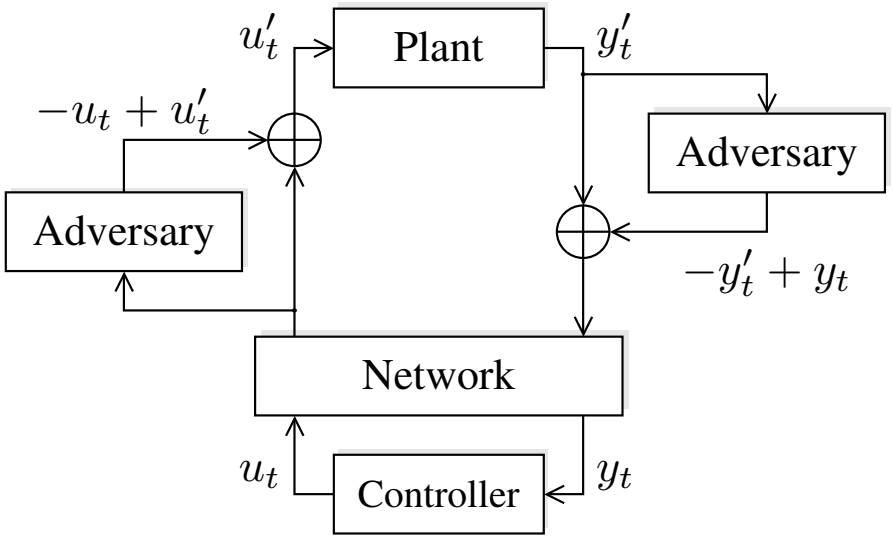

Figure 1: Representation of a cyber-physical industrial attack against a networked control system. u_t and y_t represent the correct input and output vectors of the system. u'_t and y'_t represent the attack vectors.

From a cyber perspective, the Supervisory Control and Data Acquisition (SCADA) technologies are used to control industrial environments (such as power distribution, or transportation systems). In addition, protocols based on network-wide control systems need to cover control rules such as delays and anomalies [6]. Indeed, most industrial control protocols (e.g., MODBUS, DNP3, AGA-12, PROFINET, and EtherNet/IP) are designed to provide system safety, but not information security across the network. However, there are protocols with security extensions. AGA-12 uses cryptography to add integrity and confidentiality protection, but with a high deployment cost. DNP3 can be equipped with an extension called DNP3-SA (Secure Authentication, as of the fifth IEEE-1815-2012 release), adding message integrity and authentication to DNP3. However, current cyber-physical systems use these protocols over TCP/IP or UDP/IP (e.g. MODBUS, DNP3 and PROFINET over TCP, EtherNet/IP over TCP or UDP). In this case, there are just security mechanisms up to the application layer, such as TLS and IPSec.

At the application layer, we also find protocols that have evolved. For example, PROFINET can be complemented with a new layer, PROFIsafe, which is designed to provide security, thus protecting the protocol against malfunction (e.g., transmission errors). Unfortunately, this does not provide security against intentional malicious acts [1]. It should be noted that most protocols that run over Ethernet or TCP/IP are modifications of serial protocols that do not provide security. Although the transport and network layers can provide a certain level of security, these mechanisms are not sufficient to provide protection for control data. To fully address the problem of control data protection, it is necessary to add new cyber-physical solutions to these protocols.

In the literature, some authors have proposed the use of a physical attestation at the cyber layer [28], a physical signature sent by the cyber layer to the physical layer in order to verify the correct behavior of the physical processes [24], or a signature on the physical data in order to avoid identifying the real value of the data and secure the communication [4]. In [2], Arvani *et al.* describe a detection method using discrete wavelet signal transformation. Do *et al.* [13] investigate strategies for handling cyber-physical attacks using statistical detection methods. These proposals are only valid when adversaries perform a replay or integrity attack without the ability to gain knowledge about physical processes.

1.1 Objectives and Contributions

This chapter focuses on the security between the cyber layer and the physical layer, forming cyber-physical systems. We start with a security analysis based on theoretical detection mechanisms proposed by Mo and Sinopoli [22] and Chabukswar *et al.* [10],

who study the use of stationary signatures to detect attacks to cyber-physical systems. Continuing the approach of signature-based detectors, we propose a new detection mechanism using *non-stationary signatures* to cover a larger number of threats. This new mechanism increases the attack detection rate while keeping the same performance cost as the previous approach. Next, we analyze the limitations of the new proposal. This analysis leads us to improve the detection mechanism, as well as to create a new control and security strategy capable of avoiding the security weaknesses generated by the cyber layer's membership in the physical and control domain.

Current security for cyber-physical systems focuses on either cyber adversaries or physical adversaries, but not both. For this reason, the new security challenges in cyber-physical systems require the analysis of control strategies and security mechanisms to detect attacks. This analysis will allow us to create a new control and security strategy, improving the existing detection mechanisms in the literature, in order to secure the cyber and physical layer against cyber-physical adversaries.

Contributions of the chapter: The mechanisms proposed in this chapter allow us to detect threats conducted by cyber-physical adversaries. In addition, we analyze the different cyber-physical adversaries, and we classify these adversaries according to their ability to obtain the correct behavior – or *dynamical model* – of the system. Based on this classification, we can define two different types of cyber-physical adversaries: *parametric* and *non-parametric* cyber-physical adversaries. We also address the shortcomings of centralized detection mechanisms by proposing a decentralized detection strategy that increases the robustness of the system against attacks. Then, we define a distributed detection mechanism that increases the robustness against cyber-physical attacks. Finally, we build SCADA simulations and testbeds to validate the new detection models.

This chapter is organized as follow. Section 2 provides additional preliminaries on watermark-based detector mechanisms, reports their main limitations and presents our extended approach. Section 3 presents a distributed detection mechanism, as an evolution of existing watermark-based detectors. Section 4 reports a training cyber-physical testbed to validate the mechanisms presented in this chapter. Section 5 provides some future research lines that may be undertaken. Section 6 concludes the chapter.

2 Dynamic Challenge-Response Authentication Scheme

In this section, our focus is on integrity issues due to the interconnection between the *cyber* and *physical* domains in control systems across the network. Specifically, we focus on adapting anomaly detection mechanisms, existing in the physical domain, to handle attacks as well.

The authentication scheme proposed by Mo *et al.* [21] relies on the adaptation of a real-time anomaly detector based on a *linear* and *time-invariant* model of the system. This scheme, built employing *Kalman Filters* as *linear quadratic estimators (LQE)*, and *linear-quadratic regulators (LQR)*, generates authentication signatures to protect the integrity of the physical measurements communicated across the network, because if the messages carrying these measurements are not protected, malicious actions can be taken to mislead the system. However, we show that the detection scheme proposed by Mo *et al.* only works against certain integrity attacks. We present two new models of adversaries that can evade this detector. These adversaries are classified according to the algorithm used to obtain knowledge of the system dynamics to carry out the attack. Then, we revisit the mechanism proposed in [23, 24] and evaluate its performance against the two new adversary models presented in this section. We adapt this detection scheme to handle uncovered limitations, validating the resulting approach with numerical simulations.

2.1 Problem Formulation

This chapter focuses on physical environments of industrial control systems that can be mathematically modeled as discrete linear time-invariant (LTI) systems. It is worth mentioning that a mathematical model provides a rigorous way to describe the dynamic behavior of a given system. One well-known way to describe the dynamics of such a class of systems is the state-space formulation:

$$x_{t+1} = Ax_t + Bu_t + w_t \quad (1)$$

$$y_t = Cx_t + v_t \quad (2)$$

where $x_t \in \mathbb{R}^n$ is the vector of *state variables*, $u_t \in \mathbb{R}^p$ is the control signal, $y_t \in \mathbb{R}^m$ is the output of the system, and $w_t \in \mathbb{R}^n$ and $v_t \in \mathbb{R}^m$ are the *process noise* and the *sensor measurement noise* respectively. The noises are assumed to be Gaussian white noise with zero mean and covariance Q, *i.e.* $w_t \sim N(0, Q)$ and R, *i.e.* $v_t \sim N(0, R)$. Moreover, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{m\times n}$ are respectively denoted as the state matrix, the input matrix, and the output matrix.

For the class of systems defined above, one of the most widely used control methods is the *Linear Quadratic Gaussian* (LQG) control. This control consists of two components that can be designed independently:

1. A *Kalman filter* that produces an optimal state estimate $\hat{x}_t$ of the state x_t based on the obtained noisy measurements y_t.
2. A *Linear Quadratic Regulator (LQR)* that provides the control law u_t that solves the LQR problem, based on the state estimate $\hat{x}_t$.

2.2 Detector Based on a Stationary Signature

This section briefly describes the detection scheme proposed by Mo *et al.* [23, 24]. The procedure applies to physical environments that follow a discrete-time LTI model, and are controlled by an LQG controller (cf. Section 2.1). Before presenting the detection scheme, we provide a definition of the adversary model considered in [23, 24]:

Definition 2.1 *An attacker who has the ability to listen to all messages containing the outputs of the y_t sensor, and inject messages with a y'_t signal to carry out malicious actions, is defined as a cyber adversary.*

Remark 2.1 *It is important to note that the definition given above assumes that the attacker does not possess (or attempt to gather) knowledge of the system model. For this reason, we refer to such an attacker as a cyber adversary.*

In what follows, we will call u_t^* the output of the LQR controller and u_t the control input that is sent to the physical environment (cf. Equation (1)). The idea is to superimpose on the optimal control law u_t^* a signature signal $\Delta u_t \in \mathbb{R}^p$ which serves as an authentication signal. Thus, the control input u_t is given by:

$$u_t = u_t^* + \Delta u_t \tag{3}$$

The signature signal is a random Gaussian signal with zero mean, which is independent of both w_t process noise and v_t measurement noise. This authentication signature is expected to detect the repetition and integrity attacks generated by the cyber adversary defined above. Given that the optimal control law u_t^* is equipped with the authentication signal Δu_t, a *detector* – physically co-located with the controller – can be designed with the objective of generating alarms when an attack occurs. For this purpose, Mo *et al.* [23, 24] propose to use a χ^2 detector, which is a well-known class of real-time anomaly detectors classically used for anomaly detection in control systems [7], with the objective of attack detection. Figure 2 shows the global control system equipped with the attack detector proposed in [23, 24].

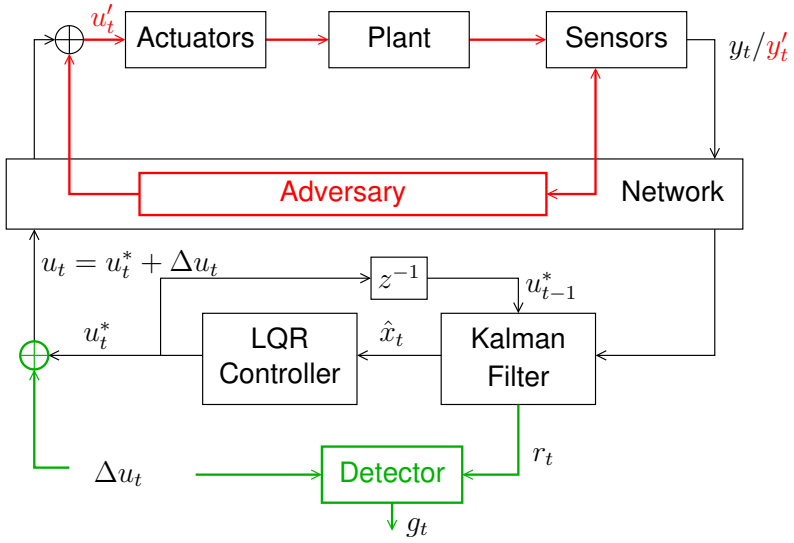

Figure 2: Signature-based protection in cyber-physical systems [24].

An *alarm signal* g_t is computed based on the residuals $r_t = y_t - C\hat{x}_{t|t-1}$ generated by the estimator. Then, g_t is compared to a threshold, γ, to decide if the system is in a normal state. The threshold is set to minimize false alarms [23, 24]. The alarm signal g_t is calculated as follows:

$$g_t = \sum_{i=t-w+1}^{t} (y_i - C\hat{x}_{i|i-1})^T \mathcal{P}^{-1} (y_i - C\hat{x}_{i|i-1}) \tag{4}$$

where w is the detection window size and $\mathcal{P}$ is the co-variance of an independent and identically distributed Gaussian input signal from the sensors.

The system is considered unattacked if $g_t < \gamma$. The system is otherwise considered as attacked and the sensor generates an alarm.

2.3 Cyber-Physical Adversaries

In this section, we introduce an improved adversary that is aware that the system uses the χ^2 detector presented above. Since the detector is based on a stationary signature signal Δu_t, we show that an adversary that is able to extract the model of the system from the control law u_t and the sensor measurement y_t, is able to perform an attack while remaining undetected.

Definition 2.2 *An attacker who, in addition to the capabilities of the cyber adversary, is able to listen to messages containing the controller's output (u_t) with the intent of improving his knowledge of the system model using a parametric or non-parametric identification model, is defined as a cyber physical adversary.*

Depending on how the system behavior is modeled, two different cyber-physical adversaries can be defined as follows:

Definition 2.3 *An attacker who uses only the previous input and output of the system to identify the system model is defined as a non-parametric cyber-physical adversary.*

Remark 2.2 *A non-parametric cyber-physical adversary may use, for example, a finite impulse response (FIR) filter-based model identification tool to identify the model of the system [31]. In Figure 2, the signals u'_t and y'_t are assumed to be the controller output and the sensor output, respectively, when an attack is occurring. We denote by $\Delta u'$ the signature estimated by the non-parametric cyber-physical adversary.*

Definition 2.4 *An attacker capable of estimating system parameters using input and output data to fool the controller's detector is defined as a parametric cyber-physical adversary.*

Remark 2.3 *A parametric cyber-physical adversary is capable of estimating system parameters using input and output data to mislead the controller's detector. This adversary can use, for example, an ARX model (autoregressive with exogenous input) or an ARMAX model (autoregressive with dynamic mean and exogenous input) to estimate the dynamics of the system [27].*

We assume that the main constraint of this adversary is the energy expended to listen and analyze the communication data, i.e., the number of samples needed to obtain the system model parameters.

2.4 Multi-Signature Based Detector

In the previous sections, we defined three types of adversaries that use different control system vulnerabilities to perform attacks ; cyber adversaries, non-parametric cyber-physical adversaries, and parametric cyber-physical adversaries. In this section, we propose a detection scheme that extends the one presented in [23, 24], to detect cyber-physical adversaries. We also study the performance loss of the new detection scheme compared to the one presented in [23, 24].

The goal of our new detection scheme is to increase the difficulty of recovering the authentication signature Δu_t from the control signal u_t, so that the probability of detecting an attack from a non-parametric cyber-physical adversary can be increased. We assume that the attacked control system uses exactly the same type of controllers and detection strategy presented in Sections 2.1 and 2.2. The only difference in the proposed detection scheme is how the signature signal, Δu_t, is generated. The control input u_t, as in the case of the detection scheme presented in Section 2.2, is computed as the superposition of the optimal control signal u_t^* produced by the LQR controller and a signal of several signatures, Δu_t. The idea is to build the authentication signature signal by alternating between N different and independent processes with different co-variance and mean. More precisely, the non-stationary signature, Δu_t, is obtained by changing periodically, with a period T, between N signals $\Delta u^{(i)}$, with $i \in \mathcal{I} = \{0, 1, \ldots, N-1\}$, extracted by different stochastic processes. Therefore, the signature signal Δu_t can be formalized as follows:

$$\Delta u_t = \Delta u_t^{(s(t,T))} \tag{5}$$

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

$$s(t,T) = \left\lfloor \frac{1}{T} \mod (t, NT) \right\rfloor \tag{6}$$

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

Using the proposed signature (cf. Equation (5)), we now have a suitable adaptive protection mechanism with two main configurable parameters; the number of distributions N and the switching frequency $f = 1/T$. Note that the original signature signal described in Section 2.1 is recovered when $frightarrow0$ and when $\Delta u_t^{(0)}$ is a Gaussian and stationary process with zero mean.

2.4.1 Validation against Non-parametric Cyber-Physical Adversaries

This section validates the previously proposed detection scheme with numerical simulations. In particular, we want to show that the proposed signature signal is able to detect non-parametric cyber-physical adversaries (cf. Section 2.3) with a higher detection rate compared to that obtained with the signature proposed in [23, 24]. The simulation is based on Matlab and Simulink models of a plant, as well as the models of the non-parametric cyber-physical adversaries. We use three different (i.e., $N = 3$) randomly switched distributions: a Gaussian distribution, a Rician distribution, and a Rayleigh distribution.

To quantify the effectiveness of the proposed detection scheme, we compute the detection rate as a function of the switching frequency. In particular, for each frequency f considered, we perform 200 Monte Carlo simulations (with randomly generated system parameters) in the case of a non-parametric cyber-physical adversary and a cyber adversary, and compute the cumulative distribution function (CDF) of the detection rate.

We first compare the performance obtained with the non-stationary signature-based detection strategy proposed in this section with that proposed in [23, 24] in the case of a cyber adversary and a non-parametric cyber-physical adversary. We consider here two switching frequencies $f_L = 0.05$Hz (change the signature after 20 steps) and $f_H = 0.14$Hz (change the signature after 7 steps). We verify that the proposed detection strategy in [23, 24], as mentioned before, can detect a cyber attack, but performs poorly when a cyber-physical adversary attacks the system. Nevertheless, the proposed detection strategy based on a non-stationary signature is able to provide a higher detection rate. In particular, we notice that the detector using higher switching frequency f_H provides better performance compared to using the lower switching frequency f_L.

Efficiently validation: Above, we validated the non-stationary signature detector using a static function $\mathcal{I}$ to define the multi-signature. Hereafter, we present the results and validations obtained for a system with the same performance loss between the detector using a stationary signature and the one using a non-stationary signature where this non-stationary signature is generated from a non-static function, $\mathcal{I}_d$. In this simulation, both detectors have a performance loss of 30%, ΔJ, compared to the optimal cost. In addition, the signature uses a dynamic function to define non-stationarity. We find that using the multi-signature (or non-stationary signature) with the same performance loss as the stationary signature, the detection ratio increases as the switching frequency varies in the range $[0, 0.14]$ Hz, where $f = 0$ is the detector of the stationary signature. We confirm that the multi-signature performance increases up to $f = 0.14$ Hz where we observe a peak before the detection ratio stabilizes. This peak before the stability indicates that $f = 0.14$ is the resonance frequency of the system. In the next section, we extend the analysis to the case of parametric cyber-physical adversaries. In addition, we test systems of different order concluding that the detection rate increases with the complexity of the system.

2.4.2 Validation against Parametric Cyber-Physical Adversaries

Previously, we have seen how the multi-signature detector is able to detect non-parametric cyber and cyber-physical adversaries. Hereafter, we extend the study to the case of parametric cyber-physical adversaries (cf. Definition 2.4). We recall that parametric cyber-physical adversaries are able to identify the parameters of the system model from the input and output signals of the plant (physical environment). A parametric cyber-physical adversary can obtain the system model with high accuracy if control commands and sensor measurements are accessible. For example, using the signature characteristic, a parametric cyber-physics adversary can use an ARX (autoregressive with exogenous input) model to define the system.

Similarly to the previous validation, we analyze the detection ratio for 200 Monte Carlo simulations using 25 order systems, against seven different parametric cyber-physical adversaries. The assumed window size is $\hat{T} = 300$. If the adversaries use a model of the system with the correct order, the detection ratio is about 8%. The set of system orders where the detection ratio does not increase drastically is $[18, 28]$. Otherwise, the probability of detecting the adversary is high. Next, we analyze the detection ratio of the same system, against a parametric cyber-physical adversary with different window sizes (125, 150, 200, 250, and 300), and with the correct system order. We conclude with the results obtained that the window size used by the adversary is inversely proportional to the detection ratio.

Remark 2.4 *A parametric cyber-physical adversary is able to obtain the system model, $H(z)$, and mislead the controller by listening to control inputs and sensor measurements. The probability of being detected is equivalent to the probability of obtaining an erroneous model. This probability is directly proportional to the order of the system; and inversely proportional to the size of the window for listening to the data channel.*

From Remark 2.4 it follows that, if we consider the real system as a black box, a misidentification of the system depends on the order of the system chosen by the adversaries to recreate the model of the system, as well as the number of samples listened to and the size of the window used by the adversaries to recompute the parameters of the target system. This can be quantified using the Mean Square Error (MSE) [20, 5]. In summary, the probability of obtaining the correct model of the targeted system is directly proportional to the order chosen by the adversaries to generate the model and inversely proportional to the number of samples recovered. The computational cost for adversaries is directly proportional to the order of the system, as such adversaries must increase the order of the model, as well as the window size in order to minimize the MSE. Therefore, the number of samples listened to before performing the attack, and the system order chosen by the adversaries are the two main parameters to escape detection.

2.5 Discussion

We have shown in this section that the detection strategy of classical stationary signature signal (single watermark approach) is not robust enough from a security perspective. In particular, we have shown that an adversary that learns about the system model is able to separate the watermark from the control signal, to evade detection and successfully attack the system. Indeed, we have shown a quantitative validation that the approach only detects *cyber adversaries*. Then, we have presented an adaptive detection scheme based on multiple signatures with two main configurable parameters: the number of distributions and the switching frequency. The main idea of the new scheme is to use multiple watermark distributions and non-stationary identification signals. The new proposal succeeds in correctly detecting non-parametric cyber-physical adversaries, under the assumption that the signature distributions change frequently. The rationale is that, the non-parametric adversary has little chances of acquiring the necessary information to acquire the watermark and bypass the detector. Moreover, we have confirmed that the multi-watermark detector approach, with the same performance loss as the single-watermark approach, has a higher detection ratio (cf. Ref. [29] and citations thereof for further details).

As we have seen, smarter cyber-physical adversaries able to dynamically adapt their behavior can successfully evade detection and disrupt systems whenever the appropriate parameters are met. A more detailed analysis of the strategy used by this new class of adversaries is detailed next. An alternative detection strategy is also proposed. The new detection strategy successfully mitigates the effects of the adversaries uncovered in our analysis.

3 Adaptive Detection Based on Control Theory

As we have shown in Section 2, the use of inadequate cyber-physical security mechanisms can have a negative effect in industrial cyber-physical systems [14, 32, 19]. These new systems need the collaboration of a very wide range of disciplines to solve the challenges in terms of autonomy, reliability, usability, functionality and cybersecurity [3]. Hereafter, we focus on the use of control theoretic solutions to detect attacks against cyber-physical systems. Traditional literature proposes the use of control strategies to maintain, for example, the closed-loop performance of the system or the safety properties of a communication network connecting the distributed components of a physical system. However, the adaptation of these strategies to manage security incidents is still a challenge.

The monitoring community is actively working on adapting traditional monitoring strategies used to detect accidental flaws and errors, towards detecting malicious attacks [30, 18, 17]. Motivated by the same goals, we present a solution that complements the signature detector to cover these weaknesses. Specifically, the new solution combines the control strategies with the challenge-response strategy analyzed and improved in the previous section. This combination allows handling integrity attacks against cyber-physical systems.

3.1 Parametric Cyber-Physical Adversaries Detection

In this section, we present a detection strategy, hereafter referred to as *Periodic and Intermittent Event-Triggered Control Watermark strategy* (PIETC-WD strategy), which aims to detect cyber and cyber-physical adversaries by complementing the strategy proposed in the previous section.

Our strategy consists of a local controller located in each sensor and a common remote controller for the whole system (distributed controller, cf. Figure 3). The cooperation between the local controllers and the remote controller allows us to create an intrusion detection policy to capture integrity attacks (cf. Definition 3.1). The local controllers manage the dynamics of the physical environment and the remote controller manages the closed loop of the system to ensure the system against integrity attacks. Note that our new system requires an additional controller for each sensor that must have enough computational power to process the data estimates to, among other things, predict the errors between the environmental data and the estimated data. The actuators do not require additional computing power. Nevertheless, during the time between two consecutive events, they must keep the last data received by the remote controller.

Such a system requires defining communication policies among sensors, actuators, and the remote controller. We define two communication policies to ensure the system: (i) *is a periodic communication policy*, with communications between the sensors and the remote controller, with period $T_{sc} = 1/f_{sc}$, and between the remote controller and the actuators, with period $T_{ca} = 1/f_{ca}$; and, (ii) *is an intermittent communication policy*, which allows data to be sent from the sensors to the remote controller if a local controller produces an alarm. Note that T_{sc} cannot be equal to T_{ca} in order to prevent intermittent communication from taking place while periodic communication is being sent.

Definition 3.1 *Periodic and Intermittent Event-Triggered Control Watermark Detector (PIETC-WD) is a detection strategy with distributed control tasks. On the one hand, the sensors monitor the system periodically, using their local controllers and signature detectors. On the other hand, the remote controller uses the estimation error received by each sensor to periodically generate the control inputs. This controller also monitors the closed loop communication with an intermittent signature.*

To execute the PIETC-WD strategy, we develop two algorithms. The first one defines the implementation of the remote controller whose input is the data sent by the sensors and the output is the set of control inputs sent to the physical system, and the alarm value. The second one shows the implementation of the local controller, placed in the sensors, whose input is the data

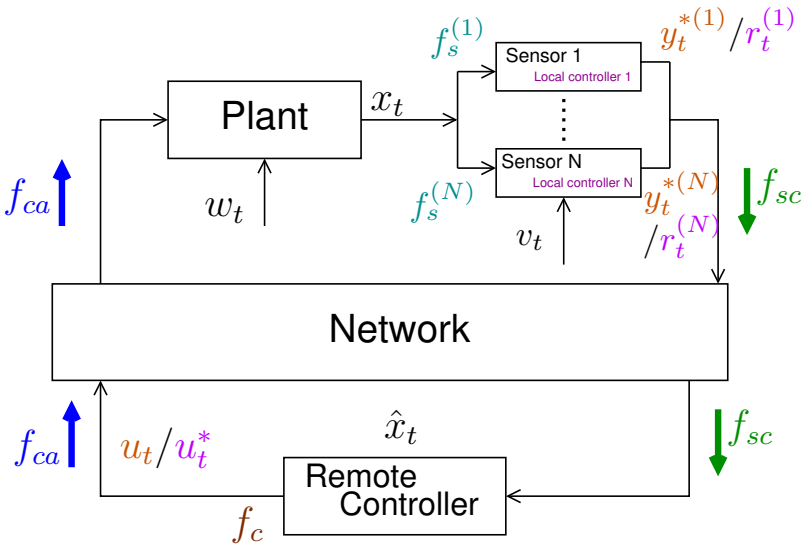

Figure 3: Diagram of a cyber-physical system with a new security strategy based on system control.

obtained from the physical system, y_t^i, with $i \in \mathcal{I} = \{0, 1, \ldots, N-1\}$, and the output is: (i) the residual of the local controllers, r_t^i (with a challenge response signature), if the alarm is not activated; or (ii) the value obtained by the physical system sensors, y_t^i, if the alarm is activated. In summary, these algorithms define how the remote controller handles data in order to increase the probability of detecting an attack, if the data sent from local controllers is not correct, or if data has been lost. Also, they determine how the sensors change the data sent to the remote controller if an alarm is activated by the sensors.

New parametric cyber-physical adversary: To validate this strategy, we introduce a new parametric cyber-physical adversary that has knowledge of the new detection strategy, in order to evaluate it. This adversary has knowledge of the new communication policies and the existence of different signatures of the data sent from the local controllers or the remote controller. However, it does not know the co-variances of the signatures, the controller parameters used to obtain the correct error between the data, or when the remote controller forces an intermittent communication.

The new adversary can detect the correlation pattern between the inputs and outputs of the physical environment. It can force intermittent communication of sensors with malisious control inputs and deceive the remote controller with read error data to obtain the pattern. Nevertheless, this adversary is not able to know when the communication is periodic or intermittent, since the attacker does not know when the controller adds, to the control inputs, the signature that generates the intermittent communication. The intermittent communication does not change the frequency of communication between the remote controller and the actuators, but produces intermittent communication between the sensors and the remote controller, which is necessary to verify the closed loop.

Using the PIETC-WD strategy, this type of adversary is detected by the localized controllers in the sensors, and by the remote controller when it checks the behavior of the closed loop. These adversaries cannot avoid the alarm in the sensors (local controllers). Nevertheless, the attackers can interrupt the communication between the sensors and the remote controller by misleading the remote controller with the correct residuals (e.g., with replayed residuals). Furthermore, in order to avoid generating an alarm in the remote controller, adversaries can switch between sending the measurement from the sensors or the residuals. However, they then have a high probability of being detected. We validate the PIETC-WD strategy against the new parametric cyber-physical adversaries in the next section.

3.2 Numerical Use Case

This section presents a practical use case where the PIETC-WD strategy proposed in the previous sections could be used in the real world. This use case is based on a chemical plant. This plant has several sensors with local controllers, actuators, and a remote controller that manages all sensor and actuator measurements. The sensors used in this use case send pressure, temperature and density information. This information is sent when an event generates an alert in a sensor, as well as periodically to indicate system behavior to the remote controller. This installation must be controlled periodically since, if the system receives incorrect or malicious control inputs capable of disturbing the system for ten consecutive periodic samples, it could reach a critical state.

To prevent an adversary from putting the system in a critical state, we use our detector strategy (PIETC-WD) with a remote controller signature management policy defined as follows:

- The controller's watermark uses a policy based on a probability to add the watermark in a specific window of samples. The windows of samples, in this use case, is assumed equal to five. For each window, the probability to add the watermark at each sample is $\beta = 50\%$. The system is able to produce $2^5 = 32$ different sequences with the same probability to be

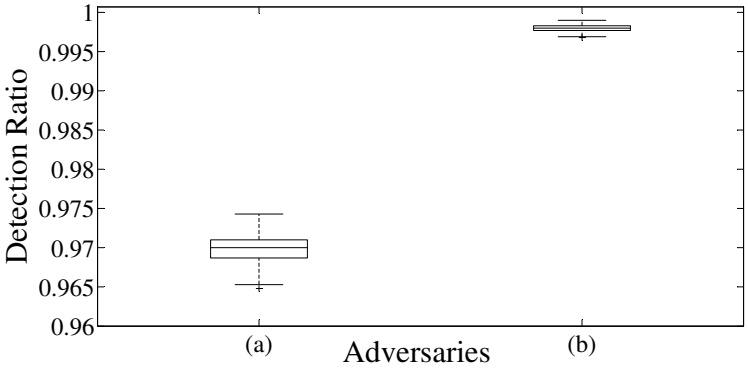

Figure 4: Detection ratio function with respect to the PIETC-WD strategy with a defined controller's watermark policy; (a) against the new parametric cyber-physical adversary; and (b) against cyber or other cyber-physical adversaries

generated, $\theta = 1/2^5$. Nevertheless, if the system send five consecutive samples without watermark, three more samples are used to add a watermark to the control input until a new control sequence starts. These three samples added to the original control sequence add $2^3 = 8$ more sequences when a window of samples has not a watermark. The three last samples have the following probability to add the watermark:

- The probability to add a watermark in the sixth sample is 60%.
- The probability to add a watermark in the seventh sample is 50% if a watermark is added in the sixth sample. Otherwise, if a watermark is not added, the probability is 60%.
- The probability to add a watermark in the eighth sample is 50%, if a watermark is added in the sixth or seventh sample. Otherwise, the probability is 60%.

Figure 4 shows the results of 200 Monte Carlo simulations using the above use case and controller signature policy (local and remote) against the cyber and cyber-physical adversary. These results show that the detection ratio is about 97% against the new parametric cyber-physical adversary and more than 99% against the other cyber and cyber-physical adversaries using the PIETC-WD strategy with a correct policy for the remote controller signature.

4 Experimental Testbed for the Detection of Cyber-Physical Attacks

Experimental testing is essential for the study and analysis of ongoing threats against cyber-physical systems. The research presented in this section addresses some actions to develop a replicable and affordable cyber-physical testbed for training and research. In this framework, our goal is to put into practice the theoretical solutions developed in the previous sections. To achieve this goal, we implement the solutions in realistic scenarios to analyze their effectiveness against intentional attacks. Specifically, we assume cyber-physical environments operated by SCADA technologies and industrial control protocols. We focus on two representative protocols widely used in industry: MODBUS and DNP3 [25, 12]. Both protocols have versions over TCP. This allows us to emulate cyber-physical environments on shared network infrastructures. We assume a Master/Slave design, which primarily dictates that slaves do not initialize any communication unless a master requests it (cf. Figure 5). One of our goals has been to combine these two protocols, both to allow flexibility and support for multiple devices with MODBUS as well as the security enhancements included in DNP3's functionality. In addition, the cyber-physical detection mechanisms based on the challenge-response strategies proposed in Section 2 are included in our SCADA testbed. Similarly, we integrate the control strategy proposed in Section 3 to experiment and analyze its actual performance. To complete the testbed, a set of attack scenarios are designed and developed to test attacks against the emulated environment. These scenarios focus on attacking the MODBUS segments of our architecture. The final goal is to analyze the effectiveness of the new security methods implemented on the emulated environment and under the application of some attack models.

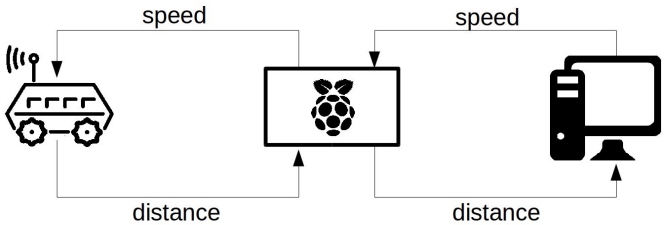

Figure 5: Test scenario overview.

4.1 Architecture Design

The proposed architecture of our cyber-physical testbed works as follows. All the elements of the system (controller, sensors and actuators), can be distributed on several nodes in a shared network combining DNP3 and MODBUS protocols (cf. Figure 6. Similarly, one or more elements can be integrated into a single device. From a software perspective, the controller never connects directly to the sensors. Instead, it is integrated into the architecture as a Programmable Logic Controller (PLC), with possible connections to other intermediate nodes. Such nodes can translate the controller commands between different protocols (e.g. MODBUS or DNP3). This architecture is capable of handling multiple industrial protocols and connecting to additional SCADA elements, such as PLCs and Remote Terminal Units (RTUs). To evolve the architecture into a complete test bench, new elements can be included in the system, such as additional RTU nodes.

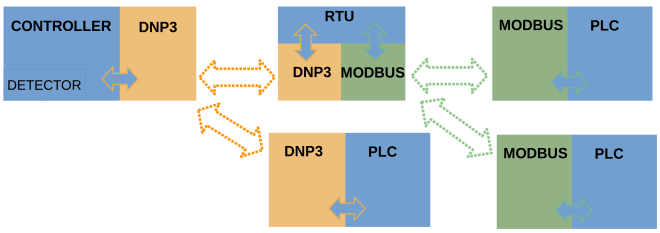

Figure 6: Abstract architecture overview.

4.2 Adversary Implementation

After getting the architecture up and running, the next requirement is to implement the adversaries reported in the previous sections. To develop these scenarios, we use a common attacker model. It implements most of the underlying capabilities of the adversaries and can be extended to implement the more specific adversaries. For example, we assume that attackers can intercept all communication exchanges between endpoints, and hence modify, store, and analyze what can be replayed to forge false data to and from communication channels. Since this is done using a testbed instead of numerical simulations, all real-life limitations are applied to the attacker. The technique *ARP poisoning* [26] is used by the attacker to intercept the channels and listen to the communications. The attacker has a passive and active mode of operation. During passive mode, the attacker only observes, processes and analyzes the data without modifying the information contained in the message payload. Ethernet header data, such as MAC addresses, are nevertheless modified due to the compromise of the ARP tables. During the active mode, the attacker starts injecting data into the hijacked communication. This injection, depending on the attacker's model, can be a replayed packet or generated by the attacker:

Replay attack: Attackers use the *ARP poisoning* technique to start listening to the connection (passive mode, from the physical layer perspective). After enough data has been recorded, the *active mode* begins. The attackers inject the old captured data. Before starting to disrupt the physical system, the attacker performs the replay attack between the sensors and the controller. Once the packets are replayed to the controller, the physical system is disrupted by falsifying data between the controller and the automata.

Injection attack: Before launching this attack, the attacker listens to connections using the passive mode *physical layer*, and analyzes the data to determine the system dynamics. This evades the signature-based authentication detector. Once the system model is deduced, the attacker starts injecting correct data into the communication channel to bypass the authentication signature. To evade the detector, the attacker calculates the effect of the signature in the system and attempts to override the detector's ability to detect changes in the return signal. Two different techniques are implemented: 1) a non-parametric adaptive filter, in order to implement the evasion technique presented in Section 2, *is a non-parametric cyber-physical attack*; and 2) autoregressive methods, such as ARX and ARMAX, in order to implement the evasion technique presented in Section 3, *is a parametric cyber-physical attack*.

The challenge of implementing these two adversaries is to synchronize the output of the adversaries when starting the attack phase. Since the target of the adversaries is to take control of the system, the data sent to the controller must be able to match the current state of the system and the correct signature correlation to avoid detection.

4.3 Attacks and Anomalies Detection

As explained in Section 2.2, the metric g_t is an operator that quantifies the difference between the output of the parametric model and the actual output of the system. An increase in g_t means that the system does not behave or respond to the signature as expected. Therefore, the system is at risk of being attacked. The value of g_t is computed for each iteration and compared to the values of some previous iterations. To eliminate false positives, we implemented an algorithm in the remote controller

to separate flaws from serious attacks or failures. Algorithm 1 is used to alert the operator only when actual intervention is required, separating flaws (e.g., latency or inaccuracy events on the sensor) from intentional attacks. For each sample received, the remote controller analyzes g_t. If g_t consecutively exceeds (more than the duration of a pre-defined *window*) a given threshold, it triggers an *alert*. Algorithm 1 is composed of four main functions, to differentiate between faults, accidents or attacks. They work as follows:

- *alarm_propagation*: this function creates a potential alarm if the detector value, g_t, is over the threshold. This alarm could be a fault peak or an attack*.
- *alert_propagation*: this function creates an alert if the *DR_value* is below a minor threshold, *min_R*, or above a maximum threshold, *max_R*. The value of *DR_value* is the difference between the detector's value at instant t, and the detector's value at instant $t-1$. The function also verifies if the system has generated faults (alerts or potential alarms), during a precise period of time (denoted by *window size*). This window is the number of samples chosen from the physical system, in order to settle the detector value, g_t. It has to be big enough to minimize the number of false positives and small enough to have enough time to react under critical situations.
- *potential_risk*: this function presents a given system risk level, taking into account alerts and potential alarms; it follows traditional qualitative risk values [16], such as (1) *slow*, (2) *medium*, (3) *high*, (4) *critical*, and (5) *very critical*.
- *real_risk*: this function warns the system under the presence of a real risk, analyzing the conditions received from the other function, i.e., number of potential alarms and alerts.

Using Algorithm 1, the detector can notify potential risks, based on qualitative impact values [16]. Along with these values, it triggers alerts to the operator whenever events are likely to be intentional attacks. Reported alerts, using window size values appropriate to the specific system, are assumed to be triggered early enough, for example, before reaching the *critical level*, to allow security operators to process the information before taking necessary countermeasures – i.e., system safety is assumed to have a higher priority than security.

4.4 Experimental Results

Using the previously defined testbed, we analyzed the detector with the stationary signature, shown in Table 1, the non-stationary signature, shown in Table 2, and the strategy defined in Section 3.1, shown in Table 3. Using the stationary signature, we can highlight that the replay attack is the most detectable scenario, with a detection rate of about 40%. The non-parametric attacker has a lower detection rate, of about 18%. This result is expected, as suggested by the theoretical conclusions and the simulation presented (cf. Section 2). The parametric attack uses the most robust system identification approach. These attacks can escape the detection process if they succeed in correctly identifying the system attributes. In terms of results, they lead to the lowest detection rate of about 12%.

	Replay Attack	**Non-parametric Attack**	**Parametric Attack**
False Negatives	64.06%	85.20%	88.63%
False Positives	0.98%	1.66%	1.35%

Table 1: Experimental results using a stationary watermark.

	Replay Attack	**Non-parametric Attack**	**Parametric Attack**
False Negatives	62.03%	54.24%	84.61%
False Positives	5.10%	3.30%	4.63%

Table 2: Experimental results using a non-stationary watermark.

We should also note that the *average detection time*, shown in Tables 4 and 5, of a replay attack is the slowest of all the analyzed scenarios. This behavior is due to the distribution properties of the signature (cf. Section 2.2). At the same time, the injection attacks (parametric or non-parametric version) are detected much faster than the repetition attack. This is due to the transition period required by the attackers to estimate the correct data before deceiving the detector. For this reason, if the attacker does not choose the precise time to launch the attack, the detector implemented at the controller is able to detect the data injected at the beginning of the attack. In addition, the attackers must also synchronize their estimates with the measurements sent by the sensors. In case the synchronization process fails, the detector identifies the uncorrelated data and reports the attack.

*Notice that we expressly use the term *alarms* to point out towards suspicious events; and *alerts* to point out to events likely to be associated to malicious attacks.

Algorithm 1 Fault and attack differentiation.

procedure DETECTOR
  $alert,\ alarm \leftarrow false$
  $potential\_alarm,\ potential\_attack \leftarrow false$
  $window \leftarrow detector\_window$
  $risk,\ potential\_risk \leftarrow 0$
alarm_propagation:
  if $detector\_value > threshold$ **then**
    $potential\_alarm \leftarrow true$
  else
    $potential\_alarm \leftarrow false$
  $old\_detector\_value \leftarrow detector\_value$
  goto *alert_propagation*
alert_propagation:
  $DR\_value \leftarrow \frac{detector\_value}{old\_detector\_value}$
  if $DR\_value < min\_R$ **or** $DR\_value > max\_R$ **then**
    $alert \leftarrow true$
    if $0 < account\_fault \leqslant window$ **then**
      $risk\_level \leftarrow risk\_level + 1$
    else
      $potential\_attack \leftarrow true$
      $alarm\_attack \leftarrow true$
  else
    $alert \leftarrow false$
  goto *potential_risk*
potential_risk:
  switch $risk\_level$ **do**
    case $\frac{window}{4}$: $potential\_risk \leftarrow potential\_risk + 1$
    case $\frac{3window}{4}$: $potential\_risk \leftarrow potential\_risk + 1$
    case $window$: $potential\_risk \leftarrow potential\_risk + 1$
  goto *real_risk*
real_risk:
  if $potential\_attack = true$ **then**
    $potential\_attack \leftarrow false$
    $alarm\_attack \leftarrow true$
    $risk \leftarrow risk + potential\_risk$
  if $alarm = true$ **or** $alert = true$ **then**
    $account\_fault \leftarrow account\_fault + 1$
  else
    $account\_fault \leftarrow 0$
  if $alarm\_attack = true$ **then** $alarm \leftarrow true$
  goto *alarm_propagation*

	Using only the watermark-based detector	**Using as well the PIETC-WD detector**
Detection Ratio	12.00%	75.25%
Average Detection Time	6.08s	6.20s
False Negatives	88.60%	38.66%
False Positives	1.35%	5.23%

Table 3: Detection performance results using the PIETC-WD detection strategy.

Regarding the results of the non-stationary signature detector shown in Table 5, we can verify that the performance obtained with this signature is compatible with the results obtained in the numerical validation (cf. Section 2). We show that the replay attack and the non-parametric attackers have a higher detection rate with this strategy, of about 60% and, 56% respectively. Then, parametric attackers have a small increase in detection rate, from 12% to 16%. Interestingly, the *average detection time* decreases compared to the stationary signature detector. Also, the number of false negatives decreases, which increases the

detection accuracy of the strategy against the implemented adversaries. However, the false positives with this strategy increase compared to the stationary signature detector. This means that, in a real testbed, the performance loss of the system is more important, as the number of false positives increases from 1.35% to 4.63%, with the same sensitivity as the previous strategy and a non-stationary signature.

Regarding the results obtained with the PIETC-WD strategy, shown in table 3, we can highlight that: (1) a system that uses only the signature-based detection mechanism against parametric attackers has a lower detection rate, about 12%. This is possible because attackers can escape the detection process if they manage to correctly identify the system attributes; and (2) a system that uses the strategy proposed in Section 3.1 has a higher detection rate, of about 75.25%. In this scenario, the detection rate increases, confirming the theoretical and simulation results reported in Section 3. The false negative ratio decreases from 88.60% to 38.66%. In terms of false positives, both scenarios show similar results, but the PIETC-WD strategy generates about 3.9% more. The time between the start of the attack and the time the attack is detected by the remote controller takes longer with the PIETC-WD strategy, since the detection signature handled by the remote controller follows a stochastic law. Therefore, we confirm that the PIETC-WD strategy increases detection performance at the expense of the time used for detection.

	Replay Attack	**Non-parametric Attack**	**Parametric Attack**
Detection Ratio	40.00%	18.00%	12.00%
Average Detection Time	10.01s	4.89s	6.08s

Table 4: Detector performance results using a stationary watermark.

	Replay Attack	**Non-parametric Attack**	**Parametric Attack**
Detection Ratio	60.00%	56.00%	16.00%
Average Detection Time	9.26s	6.27s	5.63s

Table 5: Detector performance results using a non-stationary watermark.

5 Future Directions and Research Trends

Critical infrastructures are using cyber-physical systems increasingly and massively. Their complexity also increases their vulnerability to faults and attacks. New defenses are needed, in order to complement standard detection and protection mechanisms. Intrusion and attack detection mechanisms can provide crucial components to build resilient-by-design approaches to handle extreme and complex adversaries. In this book chapter, we have addressed control-theoretic approaches capable of countering unauthorized actions from adversaries. The solutions reported in this chapter aim at identifying and attenuating the impact that adversaries may impose upon the affected elements of a cyber-physical domain. This chapter has also addressed, with limited scope, some challenges on protection in the domain with particular attention to the detection of hidden malicious actions or combined with anomalies and accidents.

In terms of perspectives for future research, several actions remain to be done. Cyber-physical systems encompass many other domains that need to be managed together to improve their resilience to attack and misuse. Perspectives include additional mechanisms allowing the quantification and assurance of resilience of cyber-physical systems, giving some steps further in achieving the security-by-design addressed in this chapter, without losing robustness of the system. Real-time needs must be addressed carefully, in order to develop new resilient systems in which adversarial attacks inflicting a security breach can be managed before detrimental events happen against the system. We envision the use of additional layers, including solutions such as Byzantine fault and intrusion tolerance techniques, as well as self-healing and diversification mechanisms.

With the aforementioned ideas in mind, a first perspective would be to include further analysis about the performance impact of the decentralized protection process presented in this chapter. Likewise, expanding the decentralized model presented in Section 3, in order to consolidate the security level of the approach, the impact of the new construction in terms of network performance, as well as the performance of the overall control system. New research in order to fully decentralize the protection strategy that has been initiated in this chapter, as well as an appropriate combination of the cyber and control-physical layers suggested in our work could be developed towards a new generation of cyber-physical SIEM (Security Information and Event Management), capable of correcting cross-layer security incidents.

6 Conclusion

This chapter is based on the premise that, in a cyber-physical system, adversaries can eavesdrop and manipulate information to disrupt the availability and integrity properties of the system. Adversaries can use techniques from both the cyber and physical layers, first to control the network layers and then to disrupt the physical devices. The combination of these techniques

can generate stealth attacks, allowing them to escape detection. Attacks against these systems can affect people and physical environments.

In terms of contributions, we started this chapter by reviewing existing technologies on cyber-physical environments from the perspective of traditional ICT security. The state of the art was completed by three main contributions: i) A first contribution was to revisit protection approaches related to stationary signatures, transforming them into an adaptive process capable of covering a larger number of adversary models; ii) We extended the resulting signature detector, used as a physical attestation in the cyber layer, by adding a decentralized strategy to extend the approach to several elements of a cyber-physical environment (not only controllers, but also sensors and actuators). The idea is to distribute the detection process over all these elements with sufficient capabilities to identify and manage the system dynamics, in order to identify malicious actions in addition to accidental flaws and errors; and iii) We validated all our proposals by integrating them into a SCADA technology testbed. The latter was implemented using SCADA protocols used in the industry (e.g., MODBUS and DNP3) and linux-based embedded devices. It allowed us to test and validate the security performance of our proposals. In addition, several adversaries capable of attacking representative scenarios were provided to complete the numerical simulations. To finish, it is worth noting that the new approach proposed in this chapter does not need synchronization among the different controllers (local or remote controllers). For this reason, it could also allow complementing other techniques such as MTD [15] whose objective is to create randomness in the cyber part of the system in order to detect and make it more resilient.

References

[1] J. Åkerberg and M. Björkman. Exploring Network Security in PROFIsafe. In *Computer Safety, Reliability, and Security: 28th International Conference, SAFECOMP 2009, Hamburg, Germany, September 15-18, 2009. Proceedings*, pages 67–80, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[2] A. Arvani and V. S. Rao. Detection and protection against intrusions on smart grid systems. *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, 3(1):38–48, 2014.

[3] R. Baheti and H. Gill. Cyber-physical systems. *The impact of control technology*, 12:161–166, 2011.

[4] P. Barbosa, A. Brito, H. Almeida, and S. Clauß. Lightweight Privacy for Smart Metering Data by Adding Noise. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14, pages 531–538, New York, NY, USA, 2014. ACM.

[5] M. Barenthin Syberg. *Complexity Issues, Validation and Input Design for Control in System Identification.* PhD thesis, KTH School of Electrical Engineering, Stockholm, Sweden, 2008.

[6] S. Brown. Functional safety of electrical/electronic/programmable electronic safety related systems. *Computing & Control Engineering Journal*, 11(11):14, February 2000.

[7] B. Brumback and M. Srinath. A chi-square test for fault-detection in Kalman filters. *IEEE Transactions on Automatic Control*, 32(6):552–554, Jun 1987.

[8] A. A. Cardenas, S. Amin, and S. Sastry. Secure control: Towards survivable cyber-physical systems. In *The 28th International Conference on Distributed Computing Systems Workshops*, pages 495–500. IEEE, June 2008.

[9] A. A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry. Challenges for securing cyber physical systems. In *Workshop on Future Directions in Cyber-Physical Systems Security*, page 7. DHS, July 2009.

[10] R. Chabukswar. *Secure Detection in Cyberphysical Control Systems.* PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, May 2014.

[11] D. Corman, V. Pillitteri, S. Tousley, M. Tehranipoor, and U. Lindqvist. NITRD Cyber-Physical Security Panel. 35th IEEE Symposium on Security and Privacy, IEEE SP 2014, San Jose, CA, USA, May 18-21.

[12] K. Curtis. A DNP3 Protocol Primer. A basic technical overview of the protocol, 2005, `http://www.dnp.org/AboutUs/DNP3%20Primer%20Rev%20A.pdf`, Last access: October 2016.

[13] V. L. Do, L. Fillatre, and I. Nikiforov. A statistical method for detecting cyber/physical attacks on SCADA systems. In *2014 IEEE Conference on Control Applications (CCA)*, pages 364–369, Juan Les Antibes, France, Oct 2014.

[14] N. Falliere, L. O. Murchu, and E. Chien. W32. stuxnet dossier. *White paper, Symantec Corp., Security Response*, 5:6, 2011.

[15] P. Griffioen, S. Weerakkody, and B. Sinopoli. A moving target defense for securing cyber-physical systems. *IEEE Transactions on Automatic Control*, 66(5):2016–2031, 2021.

[16] Group REI-cyber. La Cybersécurité des Réseaux Electriques Intelligents. White book. La Revue de l'Electricité et de l'Electronique (REE), February 2016.

[17] D. Han, Y. Mo, J. Wu, S. Weerakkody, B. Sinopoli, and L. Shi. Stochastic Event-Triggered Sensor Schedule for Remote State Estimation. *IEEE Transactions on Automatic Control*, 60(10):2661–2675, Oct 2015.

[18] W. Heemels, M. Donkers, and A. R. Teel. Periodic Event-Triggered Control for Linear Systems. *IEEE Transactions on Automatic Control*, 58(4):847–861, April 2013.

[19] J. Lee, B. Bagheri, and H.-A. Kao. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3:18 – 23, 2015.

[20] L. Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

[21] Y. Mo, R. Chabukswar, and B. Sinopoli. Detecting integrity attacks on SCADA systems. *IEEE Transactions on Control Systems Technology*, 22(4):1396–1407, July 2014.

[22] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli. Cyber-Physical Security of a Smart Grid Infrastructure. *Proceedings of the IEEE*, 100(1):195–209, Jan 2012.

[23] Y. Mo and B. Sinopoli. Secure control against replay attacks. In *Communication, Control, and Computing. 47th Annual Allerton Conference on*, pages 911–918, Monticello, IL, USA, Sept 2009. IEEE.

[24] Y. Mo, S. Weerakkody, and B. Sinopoli. Physical Authentication of Control Systems: Designing Watermarked Control Inputs to Detect Counterfeit Sensor Outputs. *IEEE Control Systems*, 35(1):93–109, February 2015.

[25] Modbus Organization. Official Modbus Specifications, 2016, `http://www.modbus.org/specs.php`, Last access: October 2016.

[26] S. Y. Nam, D. Kim, J. Kim, et al. Enhanced ARP: preventing ARP poisoning-based man-in-the-middle attacks. *IEEE communications letters*, 14(2):187–189, 2010.

[27] H. Natke. System identification: Torsten Söderström and Petre Stoica. *Automatica*, 28(5):1069–1071, 1992.

[28] T. Roth and B. McMillin. Physical Attestation in the Smart Grid for Distributed State Verification. *IEEE Transactions on Dependable and Secure Computing*, PP(99), 2016.

[29] J. Rubio-Hernan, L. De Cicco, and J. Garcia-Alfaro. On the use of Watermark-based Schemes to Detect Cyber-Physical Attacks. *EURASIP Journal on Information Security*, 2017(1):8, Jun 2017.

[30] J. Salt, V. Casanova, A. Cuenca, and R. Pizá. Sistemas de Control Basados en Red Modelado y Diseño de Estructuras de Control. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, 5(3):5–20, 2008.

[31] S. Tripathi and M. A. Ikbal. Step Size Optimization of LMS Algorithm Using Aunt Colony Optimization & Its comparison with Particle Swarm optimization Algorithm in System Identification. *International Research Journal of Engineering and Technology (IRJET)*, 2:599–605, October 2015.

[32] S. Weyer, M. Schmitt, M. Ohmer, and D. Gorecky. Towards Industry 4.0 - Standardization as the crucial challenge for highly modular, multi-vendor production systems. *IFAC-PapersOnLine*, 48(3):579 – 584, 2015.

[33] Y. Zhang, F. Xie, Y. Dong, G. Yang, and X. Zhou. High Fidelity Virtualization of Cyber-physical Systems. *International Journal of Modeling, Simulation, and Scientific Computing*, 4(2), 2013.