IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 16, 2019, accepted June 14, 2019, date of publication July 9, 2019, date of current version July 26, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927484

Mitigating Sensor Attacks Against Industrial Control Systems

LUIS F. CÓMBITA[1,3], (Student Member, IEEE), ÁLVARO A. CÁRDENAS[2], (Member, IEEE), AND NICANOR QUIJANO[3], (Senior Member, IEEE)

[1] Facultad de Ingeniería, Universidad Distrital Francisco José de Caldas, Bogotá 110231, Colombia
[2] Department of Electrical and Computer Engineering, University of California at Santa Cruz, Santa Cruz, CA 95064, USA
[3] Departamento de Ingeniería Eléctrica y Electrónica, Universidad de Los Andes, Bogotá 111711, Colombia

Corresponding author: Luis F. Cómbita (lfcombita@udistrital.edu.co; ca.luis10@uniandes.edu.co)

This work was supported in part by the Comisión de Estudios No. 015 de 2014 by the Universidad Distrital Francisco José de Caldas, in part by the Convocatoria 727 Doctorados Nacionales 2015 by Colciencias, and in part by the United States Air Force Office of Scientific Research (AFOSR) under Grant FA9550-19-1-0014 and Grant FA9550-17-1-0135.

ABSTRACT This paper describes how to design and implement a mechanism that helps to mitigate sensor attacks on industrial control systems. The proposed architecture is based on concepts from fault-tolerant control techniques. This short note explains how a Kalman filter can be used simultaneously with optimal disturbance decoupling observers to improve the performance of the mitigation mechanism for sensor attacks in cyber-physical control systems. Our proposal mitigates attacks by generating a signal that compensates the change provoked by the attacker, while at the same time reducing the number of false alarms. We demonstrate the effectiveness of our proposal using a three tanks control simulation.

INDEX TERMS Cyber-physical control systems, industrial control systems, secure control, sensor attacks mitigation.

I. INTRODUCTION

Widespread growth of new computing and network technologies has permeated industrial control systems (ICS), facilitating the pervasive use of remote sensors, and their interconnection with centralized control systems. These *cyber* infrastructures (including remote sensing and activation, digital signal processing, and computing) interact with *physical* industrial systems, creating a cyber-physical industrial control systems (CP-ICS). The goal of these CP-ICS is to improve the efficiency and reliability of these critical infrastructures; however, the inclusion of these technologies also opens the opportunity for cyber-attacks. The main purpose of these attacks is to modify the control loops to cause misbehaviors, with effects ranging from simple degradations on the performance of the control systems to those that can produce safety critical problems.

Over the years, several cyber security incidents affecting critical infrastructures have been reported [1], [2], including security problems in power plants, water treatment systems, pipelines, and transportation systems. As the threats to these systems continue to increase, the research community has been developing solutions in a variety of fields [3]–[9].

Cyber attacks are classified in two general groups [10]–[12]: i) denial-of-service (DoS); and ii) integrity attacks. The main purpose of DoS attacks is to deny access to sensor or actuator information; mathematical models for these kind of attacks are summarized in [13]. Integrity attacks are characterized by the modification of sensor and/or actuator information, compromising their integrity.

Detection and isolation of cyber-attacks in CP-ICS is a growing area of research, but on the other hand, the response and mitigation of these attacks has comparatively received less attention. Detection refers to revealing that there is an anomaly in the system caused by a cyber-attack. On the other hand, isolation focuses on identifying where the anomaly takes place (isolation is also referred as identification in some literature). The limitations of the attack monitors, and some conditions regarding the features of undetectable and unidentifiable attacks have been previously discussed [11]. In [14], the dynamical model of an irrigation channel system is used to design a bank of observers to detect and isolate attacks in the system; however, the authors do not work on control actions to mitigate the effect of an attack on the system. Integrity attacks on sensors of SCADA systems are

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Yan.

explored in [15], where the authors establish a feasibility condition for replay attacks and how to detect them with a noisy control authentication signal, but they do not discuss how to respond once an attack is identified. It is important to mention that fault-tolerant control (FTC) technology is being used as a tool for dealing with cyber-attacks [16], [17]; however, even these works that leverage the fault-tolerant literature focus on detection and isolation, mainly using unknown input observers (UIOs), they do not state how to mitigate these attacks.

Motivated by this gap in the literature, our previous work looked at the few proposals that focused on response to cyber-attacks and identified two types of responses: (i) preventive and (ii) reactive [18]. The former one focuses on the identification of vulnerabilities in a system before an attack happens, and its aim is to improve the robustness of ICS to face attacks. The latter generates a response after an alarm is triggered as a consequence of a detected attack, and its goal is to minimize the impact of the attack on the operation of the system. An increase in the resilience of ICSs through a reactive response is composed by detection and mitigation stages [19]. These stages mirror the literature in fault diagnosis, where they are known as detection, isolation, and reconfiguration [20]. Reconfiguration control actions are described as the mechanisms that trigger the response for maintaining the system stability or ensuring that the system remains in a safe zone, perhaps with some performance degradation. However, attacks and faults have significant differences, which complicates the use of reconfiguration control to face deception attacks [19]. This fact opens a gap to adapt the reconfiguration control tools in response to the distinctive features of the attacks. For instance, the adaptation of the controller in networked control systems to prevent and overcome current and future time delay switch attacks is presented in [21]. Another strategy for the attack mitigation is based on adaptive control techniques. In [22], the authors propose an adaptive controller able to deal with sensor and actuator attacks, which guarantees stability of the closed-loop dynamical system.

In this paper, we present a novel mechanism to mitigate integrity cyber-attacks on ICSs. This work shows how to produce a reactive response for the mitigation of the effect of sensor attacks on ICSs. This mechanism is validated with some simulations on a multi-input multi-output (MIMO) system testbed. We extend previous work in several ways: i) the addition of measurement noise to the readings of the sensors; ii) the inclusion of optimal disturbances decoupling observers (ODDOs), as the isolation mechanism, with the addition of a false alarms reduction mechanism; and iii) the design of a mechanism to generate a response that is able to mitigate the impact of attacks on sensors. Adding noise is a more realistic example of ICSs, and it is an extension with respect to previous works (e.g., [23]–[25]), where only noiseless scenarios have been considered. Output of ODDOs produces false alarms because usually all outputs of multivariable systems are coupled, and this causes that an attack on a unique sensor is isolated in more than one sensor. For this problem, our approach uses a novel binary logic that reduces the number of false alarms. The response mechanism consists in recovering the true information of the attacked sensor nullifying or reducing the alteration done by the attacker. The mitigation of the attack is achieved when the controller computes a trustworthy control action using the recovered information about the variables of the physical process. Several attacks have been explored, and the evaluation of our proposal is based on a key performance index such as the integral of the absolute error (IAE), which shows that lower values of IAE are obtained using our proposal instead of the conventional way that is more susceptible to false alarms.

The organization of the paper is as follows. In Section II we present the background and problem formulation describing the type of sensor attacks we consider. In Section III, we introduce a mechanism to mitigate the effect of sensor attacks, detailing the procedures to perform the detection and isolation process, the false attacks suppression process, and the control action compensation process. In Section IV we evaluate our proposal with a three-tank benchmark plant. Finally, in Section V we discuss the conclusions and future work to enhance the proposed mechanism.

II. BACKGROUND AND PROBLEM FORMULATION

An ICS provides the interconnection of equipment used to monitor and control physical equipment in industrial environments. The interconnection is based on a network that differs from the traditional enterprise networks because data is strongly linked with industrial physical processes.

Legacy industrial control systems are systems that were deployed before updated operational or security best-practices, and are not replaced because of market forces. Making the decision to keep legacy systems requires balancing the costs and risks of maintaining old systems versus the risk and expense of upgrading. However, most of legacy ICSs could work for the whole life cycle and even extend it if their security is improved. A way to achieve this purpose is to improve the security of existing legacy ICSs, with the inclusion of mechanisms to give a response and to mitigate the effect of cyber-attacks.

There are several mechanisms to improve the security of information technology systems (ITS); however, control systems cannot use the same tools for improving their security [26]. One of the different tools available in ICS (and not in ITS) is the ability to detect cyber-attacks based on the physical evolution of the state of the controlled system [27]. This evolution usually is given by differential equations for continuous-time systems and difference equations for discrete-time systems. In this work we extend this line of work for attack detection and also the attack response actions to mitigate the effect of the attack [3], [18].

A. SYSTEM SETUP

Control algorithms are commonly chosen based on the performance requirements of the controlled system. A widely used control in industry is *tracking control*, which is based on

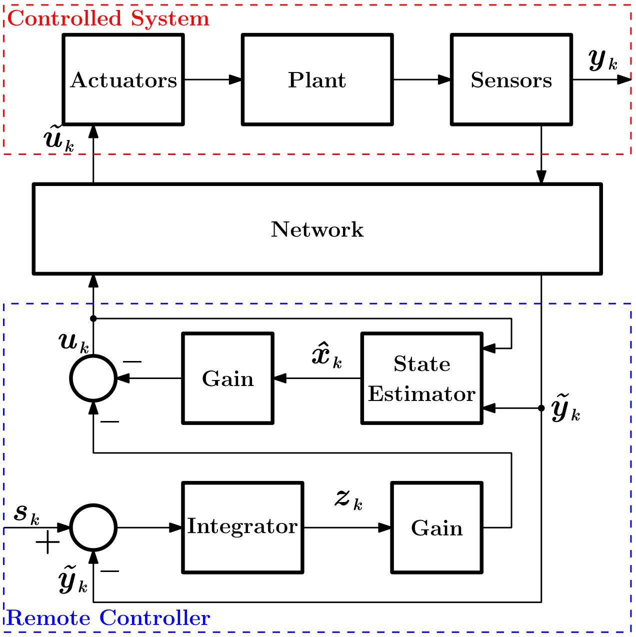

FIGURE 1. Block diagram of a typical networked tracking control system with state feedback.

state feedback together with an integral of the tracking error. Figure 1 illustrates a block diagram of a typical networked control system.

Most of the plants on ICSs are nonlinear processes, therefore nonlinear control is an important issue in industrial practice. These plants usually are modeled using state space representations, which are related with a great number of control techniques. In this work, we assume a nonlinear tracking controlled process, where the plant is modeled using a nonlinear time invariant model given by,

$$\begin{aligned} \dot{x}(t) &= f(x(t)) + g(\tilde{u}(t)) \\ y(t) &= h(x(t)), \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ represents the state of the system, $\tilde{u}(t) \in \mathbb{R}^m$ represents the control input vector after the transmission network (i.e., is the equivalent in continuous time of the signal u_k that is transmitted through the network, computed by the remote controller, and received by the actuators), and $y(t) \in \mathbb{R}^p$ represents the measurement output vector (to be transmitted).

The network used to send the signals between the controlled system and the remote controller has a random communication delay and packet dropout introduced by its limited communication capacity. Delays and cyber-attacks will affect both the control signal received by the controlled system and the sensor signal received by the remote controller, as

$$\begin{aligned} \tilde{u}_k &= \sum_{i=0}^{q} \delta(\tau_k, i) u_{k-i} + \Delta_k^a, \\ \tilde{y}_k &= \sum_{i=0}^{q} \delta(\tau_k, i) y_{k-i} + \Delta_k^s, \end{aligned} \tag{2}$$

where u_k is the control action computed by the remote controller prior to transmission, while $\tilde{u}_k$ is the control action after the process and actuators attacks are included, y_k is the signal vector from measurements of the physical variables prior to transmission, and $\tilde{y}_k$ is the sensor signal vector after including the process and sensors attacks are included. $\Delta_k^a \in \mathbb{R}^m$ and $\Delta_k^s \in \mathbb{R}^p$ represent attacks in actuators and sensors, respectively.

The Kronecker delta function $\delta(\tau_k, i)$ is used to represent the random communication delays and stochastic data missing. Delay time τ is considered as an integer number of the sampling time T_s. For the ideal case, there is no communication delay, i.e., $\tau = 0$, and only $\delta(0, 0) = 1$, hence $\tilde{u}_k = u_k$. For a communication delay time greater than zero $(1 \leq i \leq q)$, only a term of the summation is equal to 1, hence $\tilde{u}_k = u_{k-i}$. In the case that the delay produces a timeout error $q = -1$, there is no terms on the summation, and $\tilde{u}_k = 0$. Nevertheless, in this work we consider the ideal case where there is no communication delay, but we consider the case where there are cyber-attacks on the sensors.

Networked control is based on digital communication techniques, therefore, a discrete-time model for the plant is required. Typically, it is assumed that the system is operating at some nominal operation point, hence an incremental linear model (an approximation of the nonlinear plant) for the process is used in this case. In this work, the linear discrete-time invariant model of the plant is given by

$$\begin{aligned} x_{k+1} &= Ax_k + B\tilde{u}_k + B\zeta_k \\ y_k &= Cx_k + \eta_k, \end{aligned} \tag{3}$$

where $k \in \mathbb{Z}^+$ represents the discrete time instant, $x_k \in \mathbb{R}^n$ represents the state of the system, $\tilde{u}_k \in \mathbb{R}^m$ represents the control input vector after the transmission network (i.e., is the equivalent of the signal u_k that is transmitted through the network, computed by the remote controller, and received by the actuators), $y_k \in \mathbb{R}^p$ represents the measurement output vector (to be transmitted), and ζ_k and η_k are independent zero mean noise vector sequences, with covariance matrices Q and R, respectively.

The remote controller is designed to produce disturbance rejection and zero-steady state error for step inputs. For this purpose, an integrator and a state-feedback is implemented. The equation for the discrete-time integrator is given by

$$z_{k+1} = z_k + T_s(s_k - \tilde{y}_k), \tag{4}$$

where z_k is the output vector of the integrator, $s_k \in \mathbb{R}^m$ represents the reference input vector or set-point, $\tilde{y}_k \in \mathbb{R}^q$ represents the controlled output vector, and T_s represents the sampling time of the discrete-time system. The state-feedback requires the estimation of the state variables, from the available measurements. For this purpose a Kalman filter is used.

The Kalman filter provides the optimal state estimate $\hat{x}_k$. From the initial estimation of the associated error covariance matrix, $P_{k|k-1}$, the Kalman gain is computed as

$$K_k = P_{k|k-1}C^\top(CP_{k|k-1}C^\top + R)^{-1}. \tag{5}$$

After this, an update of the state estimation -using the measurement vector- and of the covariance error matrix is done

$$\begin{aligned}\hat{x}_k &= \hat{x}_{k|k-1} + K_k(\tilde{y}_k - C\hat{x}_{k|k-1}),\\ P_k &= P_{k|k-1} - K_kCP_{k|k-1}.\end{aligned} \tag{6}$$

Finally, the state estimation and the covariance error matrix is given by

$$\begin{aligned}\hat{x}_{k+1|k} &= A\hat{x}_k + Bu_k,\\ P_{k+1|k} &= AP_kA^\top + Q.\end{aligned} \tag{7}$$

As it can be noticed from (5), (6), and (7), the state estimate $\hat{x}_k$ obtained using the Kalman filter is computed using the information from all inputs u_k and all outputs $\tilde{y}_k$. The nominal control law of the system u_k is given by

$$u_k = -\begin{bmatrix}K_1 & K_2\end{bmatrix}\begin{bmatrix}\hat{x}_k\\ z_k\end{bmatrix}, \tag{8}$$

where K_1 and K_2 are vectors computed to stabilize the closed-loop control system, and to achieve the required performance. Taking into account that the Kalman filter estimation $\hat{x}_k$ is designed to converge to the state x_k, the augmented state of the whole system is $[x_k^\top \; z_k^\top]^\top$, and hence $\hat{x}_k$ in the control law (8) can be replaced by x_k.

B. CYBER-ATTACKS IN CONTROL SYSTEMS

A typical networked controlled tracking system with state feedback is depicted in Figure 1. In the ideal (non-attack) case $\tilde{y}_k = y_k$ and $\tilde{u}_k = u_k$.

When the system is under attack, equation (3) can be extended to include integrity attacks as well as DoS attacks as follows

$$x_{k+1} = Ax_k + Bu_k + B\zeta_k + \Delta_k^a, \tag{9}$$

and it is worth noticing that after the transmission, y_k becomes

$$\tilde{y}_k = Cx_k + \eta_k + \Delta_k^s. \tag{10}$$

Let us remark that attacks on sensors consist on replacing y_k (the real sensor measurement) with $\tilde{y}_k = y_k + \Delta_k^s$ (any data value output from the sensor), i.e., the attack adds what we consider as a a new input Δ_k^s to the system. Attacks on actuators consist on modifying the input of the plant (the control signal sent to the process by the controller or the programmable logic controller) adding a new input, the attack Δ_k^a. This modification affects directly the action that the actuators may execute.

Integrity attacks and faults on control systems share some similarities in that the sensor or control signals change from the real values and become less trustworthy. However, while faults are typically random and non-strategic, cyber-attacks are strategic, more deceptive, and potentially more dangerous for the safety of the system. The objective of the attacker can be economical profit, stealing private information, or causing malfunction or safety hazards in a control process. Differences between attacks and faults are significant, and, as a consequence, the existing tools of FTC cannot be used directly to detect and mitigate the effect of cyber-attacks on control systems.

In this work, deception attacks, also known as false data injection attacks, or integrity attacks, are described and discussed. For these attacks, we assume the attacker can alter the true information sent by sensors with the goal of deceiving the controller and, therefore, computing a control action that drives the control system to an unsafe or undesired behavior. These attacks can be achieved when the actual system measurements are replaced by data that are compatible with the measurement equation of the system [28], [29]. In this work, we assume the attacker knows the valid range of the measurement of sensors, then, he will produce an attack vector not trivially detectable. In [29], it is shown that an attacker can manipulate these measurements without being detected. In this attack, the attacker does not require knowledge about the model of the system, but the knowledge about current values of the measurements is enough. With current values and the span of the measurements it is easy to compute an attack vector.

The false data injection considered in this work are of the form

$$\Delta_k^s = f(k, x_k), \tag{11}$$

where $f(k, x_k)$ varies depending on the type of attack applied to the system sensors. We establish two kinds of false data injection attacks: i) *bias* attack; and ii) *static* attack.

For the *bias* attack, Δ_k^s is mathematically defined as

$$\Delta_k^s = \begin{cases} 0, & t < t_1,\\ f_i f_{s_1}, & t_1 \le k \le t_1 + 20,\\ f_i, & t_1 + 21 \le k \le t_2 - 21,\\ f_i f_{s_2}, & t_2 - 20 \le k \le t_2,\\ 0, & t > t_2, \end{cases} \tag{12}$$

where f_{s_1} and f_{s_2} are functions used to smooth the initial and final portions of the attacks, such as

$$f_{s_1} = \frac{1 + \tanh\left[(0.1(k - t_1) - 1)\pi\right]}{2} \tag{13}$$

and

$$f_{s_2} = \frac{1 - \tanh\left[(0.1(k - (t_2 - 20)) - 1)\pi\right]}{2}, \tag{14}$$

t_1 and t_2 are the initial and final times of the attack, f_i is the function that shapes the i^{th} attack itself, which is suitably defined to affect only one sensor. The purpose of smoothing functions is to produce a soft transition to the sensor data, in order to try to avoid that a detector of abrupt changes can easily detect the attack. We use (13) when the change is the addition of a positive value and, (14) is used when the change is the subtraction of a positive value.

For the *static* attack, Δ_k^s is such that $\tilde{y}_k$ becomes a static value with some noise for the duration of the attack, and it is defined as

$$\Delta_k^s = \begin{cases} -C\,x_k - \eta_k + k_{s_i} + \eta_{i_k}^s, & t_1 \leq k \leq t_2, \\ 0, & \text{otherwise}, \end{cases} \tag{15}$$

where $\eta_{i_k}^s$ is an independent zero mean noise signal, with the standard deviation equal to the sensor signal characteristics.

The fundamental feature of control systems is to maintain a set of variables with a predefined desired behavior, for instance, tracking a reference input and rejecting some disturbances. However, most control systems have not been designed to be resilient to malicious deliberate actions. Those actions aim to alter the behavior of the controlled plant in order to reach the system instability, to force some variables out of range and, in some cases, to cause harmful damage in the system and its environment.

In order to design attack-resilient systems, we need: i) to detect that an attack is taking place; ii) to isolate (identify) the attacked device; and iii) to reconfigure the system and/or change its operation to mitigate the attack (e.g., replace the sensed measurements by a virtual sensor [3]).

III. ATTACK MITIGATION APPROACH

In this section, we present a mechanism that generates a response to mitigate the effect produced by a false-data injection attack in one sensor of a control system. The mitigation response decreases the deviation in the system outputs, produced by a sensor cyber-attack. When the controller receives misleading information, it computes an incorrect control action, changing the normal operation of the control system. The attack response is based on the computation, and the posterior addition, of a correction signal to the incorrect sensor information. When the above mentioned correction is done on the tampered sensor, the controller can compute a trustworthy control action and the effect of the attack is mitigated.

To obtain trustworthy information, it is necessary to detect and isolate where the anomaly is placed. Some previous works show the use of FTC tools to detect and isolate sensor attacks in control systems [14], [19]. In this paper, we go a step further by showing a response mechanism to be used after the anomaly detection and isolation. The attack-response algorithm computes the required control action with trustworthy information to mitigate the impact that the sensor attack produces in the performance of the control system.

A. ANOMALY DETECTION AND ISOLATION

Anomaly detection and isolation algorithms have two goals i) identifying where the attack is located, i.e., which sensor information is false; and ii) identifying the time the attack is active, i.e., starting and ending time of the attack. For stochastic systems, this can be achieved using optimal filtering and robust anomaly diagnosis including unknown disturbances, which in our case are the sensor attacks. In this work, the attacks are the disturbances that we need to decouple. For each sensor of the system an observer is designed, and the optimal output estimation can be produced. To detect and isolate anomalies, the output estimation error is used as a residual which is robust against unknown disturbances and has minimal variance. A hypothesis-testing procedure is then applied to examine the likelihood of residuals, and to indicate whether or not an anomaly has occurred in any sensor of the system.

In order to detect and isolate the anomaly, an ODDO is designed for each sensor of the control system. In this work, the attacked sensors are considered as the unknown disturbance that is acting on the system. These ODDOs are used to generate the structured residuals, i.e., residuals that are insensitive to one specific disturbance, and are sensitive to the other ones. The ODDO that is insensitive to anomalies on the j^{th} sensor has as input all components of the control action vector u_k and, all but the j^{th} component of the output vector, $\tilde{y}_k^j$. An optimal state estimation of the system associated to the j^{th} sensor is obtained using an initial estimation of the associated error covariance matrix, just as it is done in the Kalman filter.

Designing the j^{th} ODDO requires a transformation to guarantee the existence of the observer and the anomaly decoupling on the j^{th} sensor, which is done by

$$\begin{aligned} H^j &= E^j(C^jE^j)^+, \\ T^j &= I_n - H^jC^j, \\ \bar{A}^j &= T^jA, \end{aligned} \tag{16}$$

where E^j is the matrix used to decouple the effect of the unknown attack on the j^{th} sensor, C^j is the resulting matrix when the j^{th} row is eliminated from the matrix C, I_n is an order n identity matrix. Then, a standard Kalman gain is calculated, similarly as in (4),

$$K1_k^j = \bar{A}^j P_{k|k-1}^j C^{j^\top} (C^j P_{k|k-1}^j C^{j^\top} + R)^{-1}. \tag{17}$$

After that, an update of the estimation of the covariance error matrix is done

$$P_k^j = P_{k|k-1}^j - K1_k^j C^j P_{k|k-1}^j \bar{A}^{j^\top}. \tag{18}$$

Some other transformation matrices, need to be updated at each iteration, and they are given by

$$\begin{aligned} F_k^j &= \bar{A}^j - K1_k^j C^j \\ K2_k^j &= F_k^j H^j \\ K_k^j &= K1_k^j + K2_k^j \\ w_{k+1}^j &= F_k^j w_k^j + T^j B u_k + K^j \tilde{y}_k^j, \end{aligned} \tag{19}$$

where $\tilde{y}_k^j$ is the vector of sensor measurements for time k, when the row correspondent to the j^{th} sensor is suppressed. Finally, the updates for the state estimated and the ahead prediction of the error covariance matrix are performed by

$$\begin{aligned} \hat{x}_{k+1}^j &= w_{k+1}^j + H^j \tilde{y}_{k+1}^j \\ P_{k+1|k}^j &= \bar{A}^j P_k^j \bar{A}^{j^\top} + T^j B Q B^\top T^{j^\top} + H^j R H^{j^\top}. \end{aligned} \tag{20}$$

The complete procedure is presented, explained, and demonstrated in [30].

Associated with the optimal output estimation from the j^{th} ODDO, the r_k^j residual vector is computed by

$$r_k^j = \tilde{y}_k - C\hat{x}_k^j \quad \text{for } j = 1, 2, \cdots, p. \tag{21}$$

The residual r_k^j, associated to the j^{th} sensor, is computed using the information from all inputs u_k and all but the j^{th} component of the outputs $\tilde{y}_k^j$. When there are no anomalies on actuators, and there is only one attack in the j^{th} sensor, the residual satisfies

$$\|r_k^j\| > \tau_I^j, \tag{22}$$

where T_{iso}^j is known as the isolation threshold for the j^{th} ODDO.

For an ideal case, the mathematical model of a system describes perfectly its behavior, the observers converge instantaneously, and hence, in absence of any anomaly on a sensor, all residuals would be always exactly equal to zero. However, for practical cases, due to modeling imperfections of the controlled system, and convergence time of observers different than zero, the residuals are not exactly zero when there are not attacks on sensors. For this reason, the threshold is determined based on reducing false isolation of attacks. The threshold determination is done based on the calculation of the residuals with no attacks on sensors of the system. From the residuals without attacks, we can define the threshold as

$$\tau_I^j = \sup_k \|r_k^j\|. \tag{23}$$

It is worth noticing that when τ_I^j is chosen to be smaller than the value in (23) some additional false anomalies are isolated, and if it is chosen to be larger than this value, some anomalies may not be detected.

The binary variable l_k^j is used to denote whether or not an attack is active at k instant on the j^{th} sensor, as

$$l_k^j = \begin{cases} 1, & \|r_k^j\| > \tau_I^j \\ 0, & \text{otherwise.} \end{cases} \tag{24}$$

However, in MIMO systems all outputs are usually coupled, and for this reason, one sensor attack in the i^{th} sensor can be wrongly isolated in another sensor, i.e., $l_k^j = 1$ for $j \neq i$. As a consequence of that, a mechanism to suppress the isolation of false attacks is presented in the next section.

B. PREVENTING ISOLATION BECAUSE OF FALSE ALERTS

The isolation mechanism described above produces imperfect results. These imperfections are the result of the coupling between all outputs of a MIMO system, i.e., an anomaly/attack on the j^{th} sensor is not just revealed on the residual of the correspond observer, but it is also revealed in the other residuals, usually delayed, and with a smaller amplitude than in the residual $\|r_k^j\|$. Hence, in this section we introduce a mechanism to correct the isolation results based on previous facts.

The false anomaly suppression is done using the previously defined assumption that establishes that only one sensor attack/anomaly can occur simultaneously, and there is no actuator attack/anomaly acting on the system. The first step of false anomalies/attacks suppression is to disable the isolation of more than one attack/anomaly simultaneously. This correction generate L_k^j variables, using their past values and the values of l_k^j:

$$\begin{aligned} L_k^j = (L_{k-1}^j \,\&\, l_k^j) || (l_k^j \,\&\, \overline{L}_{k-1}^j \,\& \\ \&\, \overline{l}_k^1 \,\& \cdots \&\, \overline{l}_k^{j-1} \,\&\, \overline{l}_k^{j+1} \,\& \cdots \&\, \overline{l}_k^p), \end{aligned} \tag{25}$$

where & represents the AND logic operator, || represents the OR logic operator, and $\bar{a}$ represents the NOT logic operator of a. Equation (25) means that L_k^j can be equal to 1, for two different situations:

1) If L_{k-1}^j and l_k^j are both equals to 1, then at the $k-1$ instant an attack was detected in the j^{th} sensor, and the attack remains active at k.
2) If l_k^j is equal to 1, the previous value of L_{k-1}^j is equal to 0, and $l_k^i = 0$, for $i \neq j$, then there is no previous attack in the j^{th} sensor, but now there is one, if and only if there is no attack in other sensors at the same time.

The second step of false anomalies/attacks suppression is related with the duration of the attack. This is done using a residual based on a Kalman filter, already designed and in use for the feedback control law calculation. This filter is used to produce an optimal estimation of the outputs when measurements are noisy. However, in this estimation, the coupling between all outputs is an advantage because the residual from Kalman filters gives accurate information about the attack/anomaly duration. Associated with the optimal state estimation obtained from the Kalman filter, one residual vector is computed as

$$r_k = \tilde{y}_k - C\hat{x}_k. \tag{26}$$

When there is no anomalies on actuators and there is one attack in any sensor, the residual obeys the next expression

$$\|r_k\| > \tau_D, \tag{27}$$

where τ_D is known as the detection threshold.

In the same way as in ODDOs, due to modeling imperfections of the controlled system and convergence time of the Kalman filter, the residual is not exactly zero when there are no attacks on sensors. The threshold determination is done based on the calculation of the residual with no attacks on sensors of the system. Hence, the supremum of $\|r_k\|$, for all k, could be chosen as the detection threshold τ_D. If a value smaller than the supremum of $\|r_k\|$, for all k, is chosen as the detection threshold some additional false anomalies are detected, and if a value greater than the supremum of $\|r_k\|$, for all k, is chosen as the isolation threshold, then some anomalies will not be detected. The binary variable d_k is used to denote whether or not an attack is active at k on any sensor of the system. In conclusion,

$$d_k = \begin{cases} 1, & \|r_k\| > \tau_D \\ 0, & \text{otherwise.} \end{cases} \tag{28}$$

The accurate information about the time duration of the attack on the j^{th} sensor is synthesized in the binary variable d_k^j, using the results from (22), (25), and (27) as

$$d_k^j = L_k^j \ \& \ d_k, \tag{29}$$

where d_k^j indicates that at the k^{th} sampling, on the j^{th} sensor, there is an attack if $d_k^j = 1$, or there is no attack if $d_k^j = 0$.

C. CONTROL ACTION COMPENSATION

Some previous works have developed similar mechanisms as the ones described above [11], [14], [19], [31]. These works have been focused on detection and isolation of cyber-attacks on control systems. In this paper we take an additional step, which consists in the addition of a mechanism with the aim of being able to mitigate the effect produced by a sensor cyber-attack of an ICS. Notice that such mechanism is added to improve the security of an existing networked controller. The proposed mechanism is developed in the same hardware where the remote controller is implemented. The purpose of this mechanism is for the system outputs to avoid having a big deviation with respect to the nominal response, when under attack.

Control action compensation is the last stage in the developed approach to mitigate the effect of a sensor cyber-attack in an ICS. It deals with the stage in which the authentic information of the sensor is recovered. The authentic information of the sensor is the signal before the attacker modifies it. In order to restore the nominal control of the system, it is necessary to find the authentic signal of the sensor using analytical redundancy. As it is explained above, the j^{th} ODDO, is designed to be insensitive to sensor attacks in sensor j, and all other ODDOs are sensitive to attacks on sensor j. For this reason, the magnitude of the attack on the j^{th} sensor is given by

$$m_k^j = C_j(\hat{x}_k^j - \hat{x}_k^i), \qquad i \neq j \tag{30}$$

where C_j is the j^{th} row of the C matrix, $\hat{x}_k^j$ is a state estimation insensitive to disturbances on j^{th} sensor and $\hat{x}_k^i$ is a state estimation sensitive to disturbances on all but the i^{th} sensor. The information given by m_k is now masked using (29) to obtain an approximation of Δ_k^s, which is subtracted from $\tilde{y}_k$ to obtain

$$\tilde{y}_k = Cx_k + \eta_k + \Delta_k^s - m_k^j d_k^j \approx y_k \tag{31}$$

where y_k is an approximation of the authentic sensor signal, nullifying the addition done by the attacker.

IV. SIMULATION RESULTS

In this section, we show some results from applying the mitigation mechanism proposed in Section III to an existing feedback control system which faces false data injection attacks. First, we describe the system and its control loop showing its normal operation behavior. Then, we describe a set of false data injection attacks. We also show the effect of the attacks on the system outputs, and we explain how the mitigation mechanism works, i.e., the obtained results step by step of the proposed approach to perform the mitigation.

A. THREE-TANKS BENCHMARK

To illustrate how FTC can be adapted and used to mitigate the effect of attacks on sensors of control systems, an existing ICS (the three tanks benchmark) is used. The nonlinear dynamics of this system are obtained using using first-principles. The approach of first-principles is based on the use of physical laws to describe the dynamic evolution of a system. In this specific case, a balance of mass is used to obtain the differential equations which are the model of the system. The model of the system is the same as the one in [32] given by

$$\begin{aligned} S\frac{d}{dt}L_1(t) &= Q_1(t) - q_{13}(t), \\ S\frac{d}{dt}L_2(t) &= Q_2(t) + q_{32}(t) - q_{20}(t), \\ S\frac{d}{dt}L_3(t) &= q_{13}(t) - q_{32}(t), \\ \frac{1}{\mu_{13}S_n}q_{13}(t) &= \operatorname{sgn}[L_1(t) - L_3(t)]\sqrt{2g|L_1(t) - L_3(t)|} \\ \frac{1}{\mu_{32}S_n}q_{32}(t) &= \operatorname{sgn}[L_3(t) - L_2(t)]\sqrt{2g|L_3(t) - L_2(t)|} \\ \frac{1}{\mu_{20}S_n}q_{20}(t) &= \sqrt{2gL_2(t)}, \end{aligned} \tag{32}$$

where the parameter values are shown in Table 1.

TABLE 1. **Parameter values of the three tank system.**

Parameter	Symbol	Value
Tank cross section area	S	0.0154 m^2
Pipe cross section area	S_n	$5 \times 10^{-5}\text{m}^2$
Outflow coefficient	$\mu_{13} = \mu_{32}$	0.5
Outflow coefficient	μ_{20}	0.6
Maximum flow rate	$Q_{i\,max}\ i \in [1, 2]$	$1.5 \times 10^{-4}\text{m}^3$/s
Maximum level	$L_{j\,max}\ j \in [1, 2, 3]$	0.62 m

The schematic diagram of the system is shown in Fig. 2. The goal of this control system is to track the liquid level of two tanks ($L_1(t)$ and $L_2(t)$) in concordance with the two set-points settled. For this case, we consider the system has three coupled tanks, with a level sensor for tanks 1 and 2 (i.e., two outputs), and two valves to regulate the intake flow in tanks 1 and 2 (i.e. two inputs). However, the state variables are the levels of the three tanks (i.e., there is no measurements in one of the three tanks).

The operation point of the system is obtained fixing the nominal intake flow as $u_1 = 3.5 \times 10^{-5}$ m^3/s and $u_2 = 3.75 \times 10^{-5}$ m^3/s. Therefore, the operation point for

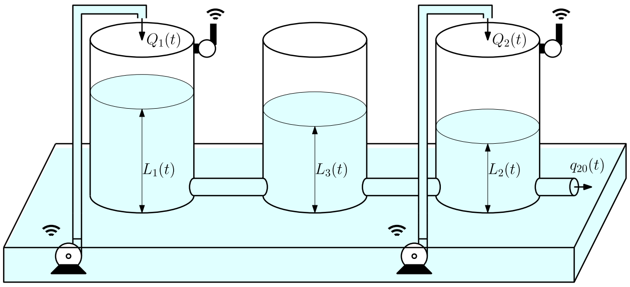

FIGURE 2. Schematic diagram of three tanks system.

the state variables of the system would be $h_1 = 0.4$ m, $h_2 = 0.2$ m, and $h_3 = 0.3$ m.

Level control of tanks on industrial scenarios is done with tracking controllers that produce disturbance rejection and zero steady state error for step inputs. The proposed control for this system given in [33] is a discrete-time controller (8) with a sampling time $T_s = 1$ s, and feedback gains given by

$$K_1 = \begin{bmatrix} 21.6 & 3 & -5 \\ 2.9 & 19 & -4 \end{bmatrix} \times 10^{-4},$$

and

$$K_2 = \begin{bmatrix} -0.95 & -0.32 \\ -0.30 & -0.91 \end{bmatrix} \times 10^{-4}.$$

In order to implement the control law, since we only have information of two level measurements, it is necessary to implement an estimator. For the open loop simulation we include white noise for the sensors, $\eta_k \sim \mathcal{N}(0,\ 5 \times 10^{-5})$, and the actuators, $\zeta_k \sim \mathcal{N}(0,\ 5 \times 10^{-6})$. Since the measurements include noise measurement, the estimation of the state variables is done using a Kalman filter. For the design of the Kalman filter a discrte-time model for the system is required. This linear model is obtained using input-output data. The data is used to estimate a discrete-time incremental linear state-space model which is an approximation of the physical nonlinear system near the operation point. The discrete-time space state model (3) is obtained using a sampling time $T_s = 1$ s as in [33], together with subspace identification techniques [34] and a similarity transformation. Therefore, the parameters of the model are given by

$$A = \begin{bmatrix} 0.9899 & 0.0005 & 0.0098 \\ 0.0004 & 0.9804 & 0.0095 \\ 0.0108 & 0.0107 & 0.9784 \end{bmatrix},$$

$$B = \begin{bmatrix} 60.1584 & 0.1660 \\ -0.3848 & 60.1895 \\ 0.4138 & 0.1935 \end{bmatrix},$$

and

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The state estimation $\hat{x}_k$ is obtained using (5), (6), and (7). Therefore, the control action can be computed now, using (8). The behavior of the closed loop system is shown in Fig. 3.

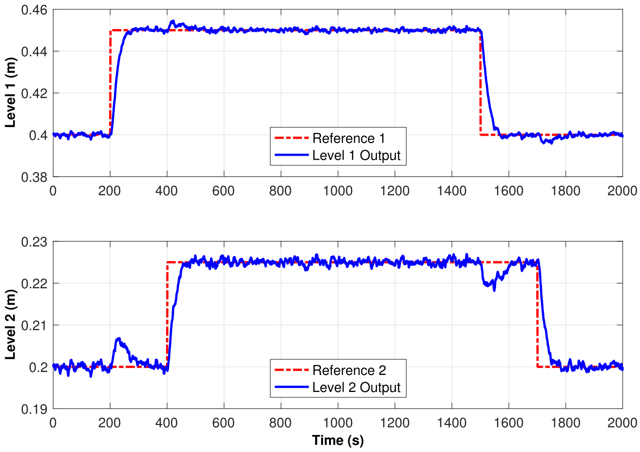

FIGURE 3. Response of closed loop control system without attacks.

There, it is seen how the control works properly for both variables $L_1(t)$ and $L_2(t)$, taking them to reach the desired value, every time the set-point varies. It can also be noticed how the system is coupled, because some small changes on the behavior appear when the references are changed.

The IAE between the response without attacks and the set-point input of the control system is used to quantify the impact of the attacks on the sensors of the system. It is important to highlight that the effect of an attack on the sensor of Level 1 has impact on the response of Level 2, and for this reason, for each attack scenario the IAE, i.e., for Levels 1 and 2 are computed. In the case there is not attacks on the control system, the IAE for Level 1 is: 3.6935, and for Level 2 is: 2.7889. These IAE values without attacks are taking into account as the reference values. Therefore, the bigger the IAE values are the bigger the impact of the attack is.

B. ATTACKS DEFINITION

In order to prove the effectiveness of the approach proposed, a set of 8 integrity attacks were applied to the system. As we mentioned before, we consider bias attacks and static attacks. Within the set of the applied attacks, there are six bias attacks, like the ones defined by (12), with their specific parameters shown in Table 2. The remaining two are static attacks, like the ones defined by (15), with their specific parameters shown in Table 3. In all of the cases, only one attack in one sensor is applied each time.

C. MITIGATION APPROACH IMPLEMENTATION RESULTS

We now evaluate our attack mitigation approach as outlined in Section III. The anomaly detection and isolation mechanisms are implemented using the existing Kalman filter used to implement the controller, and two ODDOs. The 1^{st} ODDO is designed to decouple the effect of the attacks on the sensor of tank Level 1; the estimation of the state $\hat{x}_k^1$ is obtained using (16), (17), (18), and (19); the inputs of this observer are the whole input vector u, and only the output y_2; the decoupling of attacks on the sensor for Level 1 is achieved

TABLE 2. Bias attacks applied on the system.

Attack # i	On Sensor	t_1 (s)	t_2 (s)	f_i
1	1	600	800	-20×10^{-3}
2	2	900	1100	-20×10^{-3}
3	1	900	1100	$-0.1 \times 10^{-3}(k - t_1)$
4	2	1000	1200	$-0.1 \times 10^{-3}(k - t_1)$
5	1	800	1200	$-(k - t_1) \times 10^{-4},\ t_1 \leq k \leq \frac{t_1+t_2}{2}$ $(k - t_2) \times 10^{-4},\ \frac{t_1+t_2}{2} \leq k \leq t_2$
6	2	600	1000	$-(k - t_1) \times 10^{-4},\ t_1 \leq k \leq \frac{t_1+t_2}{2}$ $(k - t_2) \times 10^{-4},\ \frac{t_1+t_2}{2} \leq k \leq t_2$

TABLE 3. Static attacks applied on the system.

Attack # i	On Sensor	t_1 (s)	t_2 (s)	k_{s_i}	η_{i_k}
7	1	800	900	0.4480	$\mathcal{N}(0,\ 5 \times 10^{-4})$
8	2	1200	1300	0.2235	$\mathcal{N}(0,\ 5 \times 10^{-4})$

TABLE 4. Key performance index comparisons of different attacks applied on system sensors.

Attack Number	Attack Kind	Sensor Attacked	IAE			Level
			w/o.R.	C.R.	N.R.	
1	B	1	7.2997 2.7904	4.6104 4.0047	4.6315 2.9526	1 2
2	B	2	3.7306 6.3866	7.1550 3.7127	3.8587 3.6059	1 2
3	B	1	5.5417 2.8479	4.4990 3.3428	4.4263 2.8644	1 2
4	B	2	3.7006 4.6440	5.0913 3.5105	3.7630 3.4056	1 2
5	B	1	6.7904 2.8494	4.3649 3.8014	4.2625 2.7832	1 2
6	B	2	3.6775 5.8342	6.5832 3.2639	3.7952 3.3311	1 2
7	S	1	6.5327 2.8991	4.0082 2.8715	3.8635 2.8262	1 2
8	S	2	3.7173 4.1333	3.7375 2.9410	3.6946 2.8781	1 2

B = Bias attack, S = Static attack, w/o.R. = Without reconfiguration, C.R. = Conventional reconfiguration, N.R. = New reconfiguration.

using the matrix $E^1 = [10^{-5} \quad 1 \quad 10^{-5}]^\top$. The design of the 2^{nd} ODDO, to decouple attacks on Level 2 sensor, is similar to the 1^{st} ODDO; in this case, the inputs of the observer are the whole input vector u, and only the output y_1; the decoupling matrix is $E^2 = [1 \quad 10^{-5} \quad 10^{-5}]^\top$.

With the state estimation from each ODDO, we can now compute the residuals r_k^1 and r_k^2. For our simulations the thresholds obtained for ODDOs are: $\tau_I^1 = 3.3 \times 10^{-3}$ and $\tau_I^2 = 1.5 \times 10^{-3}$. The threshold for residual obtained with the Kalman filter is $\tau_D = 1.5 \times 10^{-6}$.

The effectiveness of our proposal is validated using a set composed by 8 attacks. A summary of the results after applying each of the attacks 1 - 8 are shown in Table 4. The first column is utilized to specify the attack number. The Attack Kind column has two possibilities, bias attack or static attack. The sensor data measurement altered by the attacker is in the third column, named Sensor Attacked, and has two options 1 or 2, to show the corresponding level. Columns four to six show IAE values for the two outputs of the system. In these columns there are three cases, the column without reconfiguration labeled w/o.R. that shows the impact of the attack on both outputs. A column with a conventional FTC reconfiguration scheme, which exhibits the reduction of the impact of the attack utilizing FTC techniques directly, and it is labeled C.R. The last column presents the impact of the attack when the mitigation new mechanism proposed is applied, and is labeled N.R.

Results of Table 4, show that both reconfiguration mechanism reduce the impact of the attack on the output corresponding to the attacked sensor. However, in bias injection attacks the conventional reconfiguration causes a bigger impact on the opposite output, while the reconfiguration proposed maintains a better behavior in the opposite output while adjusting the attacked output. In static injection attacks, the result with both mechanisms of reconfiguration produces similar results.

1) RESULTS DISCUSSION - ATTACK #5

Now, in order to gain a better understanding of the results, we present a detailed description of two of the eight attacks utilized to show the effectiveness of our proposal. The first attack scenario analyzed is related with bias attacks. All of the attacks 1 - 6 have similar behavior, therefore without loss of generality, the attack #5 is now analyzed. The effect of the attack #5 in the outputs of the system is shown in Fig. 4. When this attack takes place, the IAE for Level 1 increases from 3.6935 to 6.7904, but the IAE for Level 2 has a little deviation from 2.7889 to 2.8494.

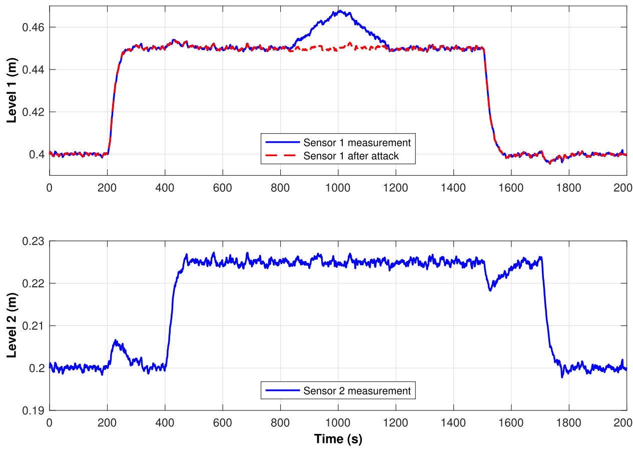

FIGURE 4. Effect of the attack #5 in the response of the control system.

The next stage on the mitigation process is the detection and isolation of the attack. This process is explained in Subsection III-A. Fig. 5 shows the result of this process.

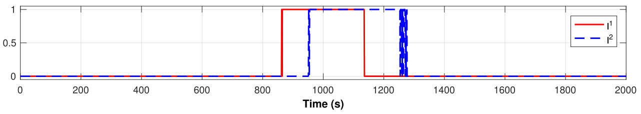

FIGURE 5. Detection and isolation of the attack #5, red line denotes isolation on Level 1, and blue line denotes isolation on Level 2.

The red continuous line shows a true detection of the attack on Level 1. Due to the soft variation at the beginning and the end of the attack, the attack is detected and isolated with some delay, between 862 s and 1136 s. However, a false isolation of an attack on Level 2 is also obtained (dashed blue line). The last is a consequence of the fact that the output named l^1 is obtained without the tampered information of the sensor of Level 1 (let us remember that the 1^{st} ODDO does not use the information of y_1), but the output l^2 is obtained using that tampered information.

The correction of the previous results of detection and isolation is done based on two facts. The first fact is that the Kalman filter, used for state feedback, is also useful to extract accurate information about the attack duration. This result is shown in Fig. 6. The second fact is that there is no simultaneous attacks on the two sensors of the system. The procedure explained in III-B is used to obtain the definitive attack isolation, and it is shown in Fig. 7.

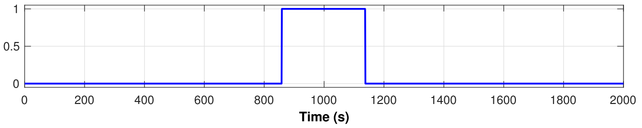

FIGURE 6. Attack duration, computed using of the Kalman filter, that is a part of the original control system, under attack #5.

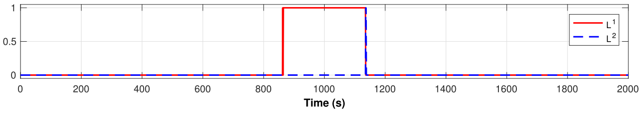

FIGURE 7. Definitive attack isolation for attack #5, red line denotes the existence and duration of an attack on the sensor of the Level 1.

A comparison between the utilization of the conventional FTC tools and the improved response obtained with our proposal is shown in the Fig. 8. It is clear that the main problem of the conventional FTC method is the degradation of the output of Level 2, when the attack on the sensor for Level 1 is mitigated. Using IAE values to compare the system behavior for the Level 1, without reconfiguration mechanism the value is 6.7904, with the conventional mechanism the value is 4.3649 and, with the improved mechanism it is 4.2625. In the same way, IAE the value for Level 2 without reconfiguration is 2.8494, with reconfiguration using the conventional FTC tools the value is 3.8014, and with our proposal it is 2.8262. IAE values, and visual inspection of Level 2 in Fig 8, show that the approach proposed keep the behavior of Level 1, where the attack takes place, and improves the behavior of Level 2, where the conventional reconfiguration process affects the system in a negative way.

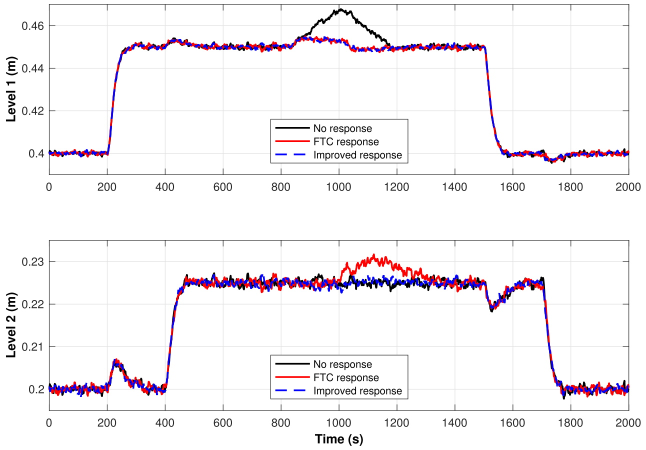

FIGURE 8. Mitigation response to sensor of Level 1 attack #5 without mitigation response and with two different mechanisms of reconfiguration.

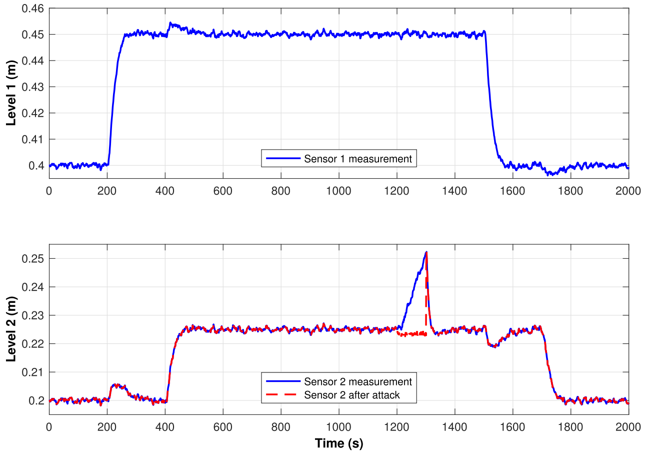

FIGURE 9. Effect of the attack #8 in the response of the control system.

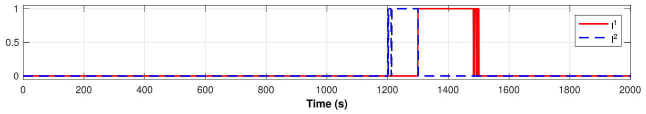

FIGURE 10. Detection and isolation of the attack #8, red line denotes isolation on Level 1, and blue line denotes isolation Level 2.

2) RESULTS DISCUSSION - ATTACK #8

Similarly to the first scenario, attacks 7 and 8 have a similar behavior, and the second attack scenario corresponds to the detailed analysis of attack #8. The effect of the attack #8 in the outputs of the system is shown in Fig. 9, and the results of attack detection and isolation are shown in Fig. 10. The attack duration and the attack isolation are shown in Figs. 11 and 12. Finally, the results of the two mechanisms of reconfiguration are shown in Fig. 13. Attack #8 also has effect on the IAE index; for Level 1 (sensor without attack) there is a little

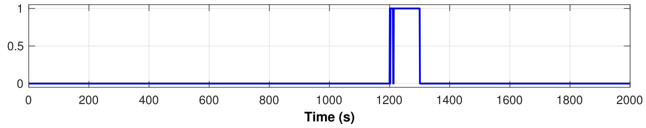

FIGURE 11. Attack duration, computed using of the Kalman filter, that is a part of the original control system, under attack #8.

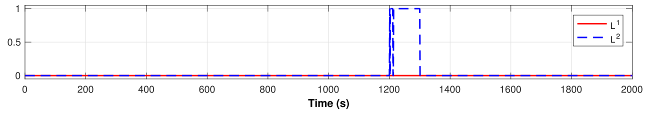

FIGURE 12. Definitive attack isolation for attack #8, red line denotes the existence and duration of an attack on the sensor of the Level 1.

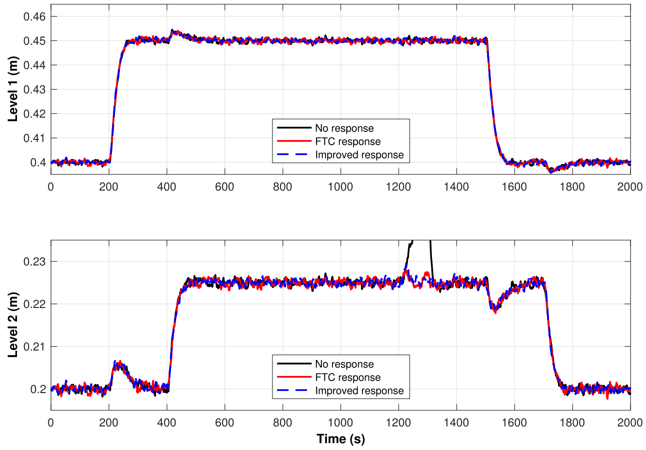

FIGURE 13. Mitigation response to sensor of Level 1 attack #5 without mitigation response and with two different mechanisms of reconfiguration.

increment from 3.6935 to 3.7173, and for Level 2 (sensor under attack) there is an increment from 2.7889 to 4.1333 (see Fig. 9).

It is worthwhile to mention that the effect of attack #8 is greater than the effect of attack #5, in the sense that the amplitude deviation accomplished with attack #8 is higher that the one with attack #5. Also, due to the integrators in the controller, the maximum magnitude of attacks like #7 and #8 is proportional to the attack duration. It is important to emphasize that attack #8 requires less resources for the attacker than attack #5, i.e., in attack #5 the attacker needs to know the current value of the sensor to add a value and then send the resultant value to the controller; but, in the case of attack #8, the attacker only needs to know the desired value of the system system, and always sends the same distortion for all the duration of the attack.

Due to the noise included in the attack, as in (15), the isolation obtained exhibits an intermittency at the beginning of the attack between 1200 s and 1213 s. However, the end of the attack on Level 2 is well detected at 1300 s. (See Fig. 10).

Unlike the results obtained for attack #5, with the reconfiguration in the case of attack #8, the effect of the conventional reconfiguration on the opposite output is insignificant. The IAE value in Level 1 for this attack, without reconfiguration mechanism is 3.7173, with the conventional mechanism is 3.7375, and with the improved mechanism it is 3.6946. These IAE values show that in this case there are no significant deviations. Now, an examination of the IAE values for Level 2, without reconfiguration is 4.1333, with reconfiguration using the conventional FTC tools is 2.9410, and with our proposal is 2.8781. It is clear, that both reconfiguration mechanisms produce similar results on the outputs of the systems.

V. CONCLUSIONS AND FUTURE WORK

In this article, we propose a novel mechanism to achieve mitigation of integrity attack effects in ICSs. The proposed mechanism is based on FTC. One of the interesting characteristics of the proposed mechanism is that it is easy to implement in any system that uses a digital control, since every stage of it (detection and isolation, false attacks isolation suppression, and control action compensation) can be computed at each sampling time.

We show the usefulness of the proposed mechanism in a system (a benchmark system designed to prove FTC techniques) facing integrity attacks (false data injection attacks), see Figs. 8 and 13. However, results for bias attacks are better than the ones for static attacks, mainly because the incidence of the attack in the opposite variables. That is, given a working control system, bias attacks always affect the attacked output as well as the other systems outputs with some delay, while static attacks only affect the attacked output.

Related to future work, we plan to test our mechanism for other kinds of attacks. Similarly, the proposal can be extended to be able to mitigate the effect of attacks on actuators (not only sensors). We also are interested in utilizing some statistics, which have been proven to work in FTC, in order to see if the current results would be comparable or could even improve the proposed architecture.

REFERENCES

[1] R. J. Turk, "Cyber incidents involving control systems," Idaho Nat. Lab., Idaho Falls, ID, USA, Tech. Rep. 10, 2005.

[2] B. Miller and D. C. Rowe, "A survey of scada and critical infrastructure incidents," in *Proc. 1st Annu. Conf. Res. Inf. Technol.*, 2012, pp. 51–56.

[3] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proc. 6th ACM Symp. Inf., Comput. Commun. Secur.*, 2011, pp. 355–366.

[4] F. Hu, Y. Lu, A. V. Vasilakos, Q. Hao, R. Ma, Y. Patil, T. Zhang, J. Lu, X. Li, and N. N. Xiong, "Robust cyber–physical systems: Concept, models, and implementation," *Future Gener. Comput. Syst.*, vol. 56, pp. 449–475, Mar. 2016.

[5] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—A survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.

[6] M. Wolf and D. Serpanos, "Safety and security in cyber-physical systems and Internet-of-Things systems," *Proc. IEEE*, vol. 106, no. 1, pp. 9–20, Jan. 2018.

[7] J. Giraldo, E. Sarkar, A. A. Cárdenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in cyber-physical systems: A survey of surveys," *IEEE Design Test*, vol. 34, no. 4, pp. 7–17, Aug. 2017.

[8] R. Altawy and A. M. Youssef, "Security tradeoffs in cyber physical systems: A case study survey on implantable medical devices," *IEEE Access*, vol. 4, pp. 959–979, 2016.

[9] H. He and J. Yan, "Cyber-physical attacks and defences in the smart grid: A survey," *IET Cyber-Phys. Syst., Theory Appl.*, vol. 1, no. 1, pp. 13–27, Dec. 2016.

[10] A. A. Cárdenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *Proc. 28th Int. Conf. Distrib. Comput. Syst. Workshops*, Jun. 2008, pp. 495–500.

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

[12] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *Proc. IEEE Amer. Control Conf. (ACC)*, Jun. 2013, pp. 3344–3349.

[13] V. L. Do, "Sequential detection and isolation of cyber-physical attacks on SCADA systems," Ph.D. dissertation, Univ. Technol. Troyes, Troyes, France, 2005.

[14] S. Amin, X. Litrico, S. S. Sastry, and A. M. Bayen, "Cyber security of water SCADA systems—Part II: Attack detection using enhanced hydrodynamic models," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 5, pp. 1679–1693, Sep. 2013.

[15] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Sep./Oct. 2009, pp. 911–918.

[16] Y. Li, J. Li, X. Luo, X. Wang, and X. Guan, "Cyber attack detection and isolation for smart grids via unknown input observer," in *Proc. 37th Chin. Control Conf. (CCC)*, Jul. 2018, pp. 6207–6212.

[17] A. Ameli, A. Hooshyar, E. F. El-Saadany, and A. M. Youssef, "Attack detection and identification for automatic generation control systems," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4760–4774, Sep. 2018.

[18] L. F. Cómbita, J. Giraldo, A. A. Cárdenas, and N. Quijano, "Response and reconfiguration of cyber-physical control systems: A survey," in *Proc. IEEE 2nd Colombian Conf. Autom. Control (CCAC)*, Oct. 2015, pp. 1–6.

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

[20] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 3, pp. 636–653, May 2010.

[21] A. Sargolzaei, K. K. Yen, M. N. Abdelghani, S. Sargolzaei, and B. Carbunar, "Resilient design of networked control systems under time delay switch attacks, application in smart grid," *IEEE Access*, vol. 5, pp. 15901–15912, 2017.

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

[23] L. F. Cómbita, A. A. Cárdenas, and N. Quijano, "Mitigation of sensor attacks on legacy industrial control systems," in *Proc. IEEE 3rd Colombian Conf. Autom. Control (CCAC)*, Oct. 2017, pp. 1–6.

[24] L. F. Combita, J. A. Giraldo, A. A. Cárdenas, and N. Quijano, "DDDAS for attack detection and isolation of control systems," *Handbook of Dynamic Data Driven Applications Systems*. Cham, Switzerland: Springer, 2018, pp. 407–422.

[25] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1454–1467, Jun. 2014.

[26] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *Proc. 3rd Conf. Hot Topics Secur. (HOTSEC)*, Berkeley, CA, USA, 2008, p. 6:1–6:6.

[27] D. I. Urbina, J. A. Giraldo, A. A. Cárdenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, "Limiting the impact of stealthy attacks on industrial control systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2016, pp. 1092–1105.

[28] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. IEEE 1st Int. Conf. Smart Grid Commun.*, Oct. 2010, pp. 214–219.

[29] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. Comput. Commun. Secur.*, 2009, pp. 21–32.

[30] J. Chen and R. J. Patton, *Robust Model-based Fault Diagnosis for Dynamic Systems*. Norwell, MA, USA: Kluwer, 1999.

[31] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1396–1407, Jul. 2014.

[32] *Laboratory Setup: Three-Tank System DTS200*. Duisburg, Germany: Amira GmbH, 2002.

[33] H. Noura, D. Theilliol, J.-C. Ponsart, and A. Chamseddine, *Fault-Tolerant Control Systems: Design and Practical Applications*. New York, NY, USA: Springer, 2009.

[34] P. van Overschee and B. de Moor, *Subspace Identification for Linear Systems: Theory, Implementation, Applications*. New York, NY, USA: Springer, 1996.

LUIS F. CÓMBITA received the B.S. degree in electronics engineering from Universidad Distrital, Bogotá, Colombia, in 1992, and the M.S. degree in electrical engineering from the Universidad de los Andes, Bogotá, Colombia, in 2002, where he is currently pursuing the Ph.D. degree.

He joined the Engineering Faculty, Universidad Distrital Francisco José de Caldas, Bogotá, as an Auxiliar Professor, in 1997, where he is currently an Assistant Professor. His current research interests include cyber-physical systems security, modelling and simulation of dynamical systems, and industrial control systems.

ÁLVARO A. CÁRDENAS received the B.S. degree from the Universidad de Los Andes, Colombia, and the M.S. and Ph.D. degrees from the University of Maryland at College Park, College Park. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, University of California at Santa Cruz, Santa Cruz. His research interests include cyber-physical systems and the IoT security and privacy, network intrusion detection, and wireless networks.

NICANOR QUIJANO received the B.S. degree in electronics engineering from Pontificia Universidad Javeriana, Bogotá, Colombia, in 1999, and the M.S. and Ph.D. degrees in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2002 and 2006, respectively.

He joined the Electrical and Electronics Engineering Department, Universidad de los Andes (UAndes), Bogotá, as an Assistant Professor, in 2007, where he is currently a Full Professor and the Director of the Research Group in control and automation systems. His current research interests include hierarchical and distributed optimization methods using bio-inspired, and game-theoretical techniques for dynamic resource allocation problems, especially those in energy, water, and transportation.