Cyberattack and Machine-Induced Fault Detection and Isolation Methodologies for Cyber-Physical Systems

Mahdi Taheri, Khashayar Khorasani, Iman Shames, and Nader Meskin

Abstract—In this article, the problem of simultaneous cyberattack and fault detection and isolation (CAFDI) for both centralized and large-scale interconnected cyber-physical systems (CPSs) is studied. The proposed methodologies include centralized and distributed CAFDI approaches, which involve the use of two filters on the plant and command and control (C&C) sides of the CPS, as well as an unknown input observer (UIO)-based detector on the plant side. The article characterizes the conditions under which the proposed methodologies can detect various types of deception attacks, such as covert attacks, zero dynamics attacks, and replay attacks. In the proposed centralized CAFDI methodology, the transmission of estimates from the C&C side filter to the plant side is required, with the assumption that a certain number of communication channels are secured. Consequently, a bank UIO-based detectors are utilized on the plant side to detect and isolate anomalies. It is also assumed that adversaries have knowledge of system parameters, filters, and the UIO-based detector. To address the limitations of secure communication channels, modifications to the two side filters and the UIO-based detector have been developed and implemented that eliminates the need for any secured communication channel in the modified CAFDI module. However, information must now be sent to and received from the plant side filter. Consequently, we develop a distributed CAFDI methodology for the interconnected large-scale CPS which consists of several subsystems. Finally, a hardware-in-the-loop (HIL) simulation of a four-area power network system under presence of both cyberattacks and faults using an OPAL-RT real-time simulator and Raspberry Pi is provided to illustrate the effectiveness of our proposed distributed CAFDI methodology.

Index Terms—Actuator attack, covert attack, cyberattack and fault detection and isolation (CAFDI), cyber-physical systems (CPS), sensor attack, unknown input observer (UIO), zero dynamics attack.

I. INTRODUCTION

CYBER-PHYSICAL systems (CPSs) are monitored and controlled by distributed sensors, actuators, and embedded computers that are interconnected via communication networks [1]. Our today’s life massively depends on the CPS due to their wide range of applications in various areas, such as power systems and smart grid, next generation aerospace and transportation systems, and process control and water treatment networks, among others [2]. Through employing CPS for these applications one can provide unique capabilities to accomplish high-level goals and reliable performance of complex tasks [3].

Anomalies and machine-induced faults as well as malicious cyberattacks in physical components of the CPS do occur and are observed in actuators and sensors. In recent years, cyber security challenges in CPS, that include cyberattacks on communication networks, have attracted significant interest [2], [4], [5], [6]. Nevertheless, the problem of simultaneous diagnosis of cyberattacks and faults has not been fully addressed in the literature.

A special type of cyberattack is defined as the deception attack in which an adversary changes the transmitted information of the system’s input or output by compromising the CPS network communication channels. This article studies the cyberattack and fault detection and isolation (CAFDI) problem of CPS in presence of machine-induced faults as well as malicious deception cyberattacks, such as covert attacks, zero dynamics attacks, replay attacks, and false data injection attacks. Covert attacks and zero dynamics attacks are defined as undetectable attacks [7], [8], [9], [10], since they have no impact on the received output measurements on the command and control (C&C) side of the CPS.

Furthermore, we consider both centralized and large-scale interconnected CPS, which leads us to develop and propose centralized and distributed CAFDI methodologies. In certain applications, such as a single unmanned aerial vehicle (UAV), due to the centralized architecture of the system, having a centralized CAFDI monitoring system is desirable, whereas
in large-scale and geographically dispersed CPS, such as power networks and smart grids, using a distributed CAFDI methodology is more suitable and practical.

In the literature, fault detection methods have been utilized to detect cyberattacks. However, due to the inherent differences between cyberattacks and machine-induced faults, in some cases, such as covert attacks, zero dynamics attacks, and replay attacks’ fault detection methods fail to detect cyberattacks. This is due to the fact that faults represent structural physical anomalies in the system, whereas cyberattacks are injected intentionally by an intelligent adversary with the purpose of damaging the nominal behavior of the system without being detected. Consequently, conventional fault diagnosis algorithms should be fundamentally generalized to accommodate the malicious intelligent adversary cyberattacks threats.

As a brief overview, the geometric-based fault detection methodologies were proposed in [11] and [12] to obtain necessary and sufficient conditions for existence of observers that can be used to generate a residual signal for the purpose of fault detection and isolation (FDI). In addition to geometric approaches, many algebraic model-based FDI methods have been introduced in the literature, such as unknown input observer (UIO) [13], [14], interacting multiple model [15], multiple model [16], distributed detection algorithms [17], [18], and parity equation-based approaches [19], [20].

For the cyberattack detection problem, a periodic modulation scheme with the idea of changing the behavior of the control input was proposed in [21] to detect covert and zero dynamics attacks in the CPS. A distributed methodology that utilizes two local observers to detect covert attacks in large-scale interconnected systems was proposed in [22]. A methodology that generates watermarked control inputs have been proposed in [23] that can be employed to detect replay attacks in the CPS.

To detect stealthy cyberattacks on output measurements in the CPS such as replay attacks, a bank of multiplicative sensor watermarking filters was proposed in [10]. In [24], geometric theory was used to define zero dynamics attacks and show their impact on the system. A method was also proposed to add perturbations to the system matrices of the system (A, B, C) to change the zero dynamics of the system so that the adversary can no longer excite these new zero dynamics modes. In [25], a sensor coding method was proposed that reveals stealthy false data injection attacks by changing the direction of cyberattacks where an algorithm to compute the coding matrices was designed, and finally, a time-varying coding approach was developed for the case when the adversary is capable of estimating a static coding matrix.

Weerakkody and Sinopoli [26] developed a moving target approach in which certain time-varying external dynamics are added to the system. Leveraging the moving target approach, the extended dynamics of the system become unknown to adversaries and they no longer are capable of executing covert attacks and replay attacks. In [27], the state-space model of a linear system was augmented by adding switching auxiliary dynamics that are unknown to the adversary and a switched Luenberger observer was designed to detect covert and zero dynamics attacks, however, for implementation purposes the extended system and the switched observer need to be synchronized.

In all the above work, the problem of detection and isolation of cyberattacks and faults has not been addressed. Hence, using the above methods, a fault in the system can misleadingly be detected as a cyberattack and vice versa. Consequently, if a fault is misleadingly detected as a cyberattack, the CPS operators may consider a certain course of action, and therefore countermeasure strategies that are designed to cope with the cyberattack threats will not resolve the machine-induced fault problem and will not recover the CPS. On the other hand, if a cyberattack is misleadingly detected as a fault, the CPS operators cannot resolve the problem by utilizing fault-tolerant control methodologies or by repairing components of the CPS that are misleadingly diagnosed to be defective. This issue has been taken into consideration in [28] where a methodology to detect and isolate faults and cyberattacks has been suggested. An event-triggered adaptive estimator is designed and proposed in [28] which can be used to isolate sensor replay attacks and sensor faults. In [29], Bayesian network models have been utilized and constructed to distinguish between cyberattacks and faults in sensor measurements for floodgates in water management infrastructures.

Due to stealthiness of covert and zero dynamics attacks, it is of paramount importance to develop methodologies that can be used to detect and isolate them. However, in [28] and [29], undetectable cyberattacks such as covert attacks and zero dynamics attacks have not been considered. In addition, due to existence of physical component faults in the CPS, one needs to also clearly detect and isolate actuator and sensor faults and cyberattacks on actuators and sensors. Moreover, in [28] and [29], only detection and isolation of sensor faults and cyberattacks on sensor measurements have been investigated. Hence, this article aims at addressing the problem of CAFDI for the CPS that are under cyberattacks such as covert attacks, zero dynamics attacks, and replay attacks under both centralized and distributed architectures.

In our proposed centralized and distributed CAFDI methodologies, two filters are designed on both the plant side and the C&C side of the CPS that are interconnected via communication channels that can be compromised by the adversary. Moreover, on the plant side, UIO-based detectors are designed to generate residuals for detecting and isolating actuator cyberattacks, sensor cyberattacks, as well as actuator faults, and sensor faults while the adversary have a complete knowledge of the filters and the UIO-based detectors. Several types of cyberattacks which in the literature are referred to as detectable and undetectable (see [2], [9], [10]) can be detected using our proposed methodology. However, in the proposed centralized methodology, it is assumed that the adversary does not have access to all the communication channels among the filters. Consequently, to relax our assumption on the number of secure communication channels, the two side filters and the UIO-based detector are modified, where without having any secure communication channel one is capable of detecting and isolating cyberattacks and faults. However, the modified CAFDI filters require additional communication channels which can be non-secure.

In the proposed distributed CAFDI methodology for large-scale interconnected CPS, the UIO-based detector of each subsystem communicates its state estimations with the UIO-based detectors of nearby subsystems. Hence, each sub-
system can detect and isolate its cyberattacks and faults as well as anomalies in its neighboring subsystems. However, it is assumed that there exists one secure communication channel from the C&C side filter to the plant side filter and one secure communication channel form the plant side filter to the C&C side filter. It should be noted that having one secure communication channel implies that from the transmitted vector of the C&C side and plant side filters states, at least one element is not manipulated by adversaries. We also study the performance of our proposed distributed CAFDI methodology when there exist no secure communication channels. We show that having secure communication channels is a necessary condition for isolating actuator and sensor cyberattacks from cyberattacks on the communication channels among both side filters. By utilizing both filters and detectors, we propose and derive conditions under which an adversary that performs cyberattack on the communication channels cannot eliminate the impacts of actuator and sensor cyberattacks.

To summarize, the main contributions of this article are stated as follows.

1) Based on both the plant side and the C&C side centralized estimation and observation methodology, design conditions are developed and provided that can be used to detect and isolate actuator cyberattacks, sensor cyberattacks, actuator faults, and sensor faults in a centralized architecture.

2) A distributed filter design methodology based on observing the system from both the plant side and the C&C side is introduced and developed that can be utilized to detect and isolate both cyberattacks and machine-induced faults in large-scale interconnected systems.

3) By utilizing our proposed methodologies, cyberattacks such as covert attacks, zero dynamics attacks, replay attacks, and false data injection attacks can be detected and isolated.

The remainder of the article is organized as follows. Mathematical models of the systems that take into account faults and cyberattacks and the definition of undetectable attacks are provided in Section II. In Section III, our proposed centralized CAFDI methodology that consists of two side filters, the UIO-based detector and residual signals are developed and investigated. Design conditions for the distributed CAFDI methodology are proposed and developed in Section IV. To illustrate and demonstrate the effectiveness and capabilities of our analytical results, a hardware-in-the-loop (HIL) simulation environment and a case study is presented and extensively investigated in Section V. Conclusions are provided in Section VI.

II. PROBLEM STATEMENT AND FORMULATION

A. Cyber-Physical Systems Model

In this article, a strictly proper linear time-invariant (LTI) CPS of the form given below is studied

\[
\dot{x}_p(t) = A^s x_p(t) + B^s u(t) + L_1 f_1(t) + N^s \omega^s(t)
\]

\[
y_p(t) = C^s x_p(t) + L_2 f_2^s(t) + v^s(t)
\]

where \(x_p(t) \in \mathbb{R}^n\) represents the state, \(y_p(t) \in \mathbb{R}^p\) denotes the measured output on the plant side, \(u^s(t) \in \mathbb{R}^m\) denotes the control input, \(f_1(t) \in \mathbb{R}^{m_t}\) and \(f_2^s(t) \in \mathbb{R}^{p_t}\) correspond to actuator and sensor faults, respectively. Moreover, \(\omega^s(t) \in \mathbb{R}^n\) and \(v^s(t) \in \mathbb{R}^p\) denote zero mean wide-sense stationary (WSS) random Gaussian processes that represent process and measurement noise with the covariance matrices \(Q\) and \(R\), respectively. The quadruple \((A^s, C^s, B^s, N^s)\) has appropriate dimensions and describe the CPS characteristics, and the known pair \((L_1, L_2)\) capture the fault signatures.

In case of injection of a cyberattack on actuators, the control input is expressed and changed to

\[
u^s(t) = u(t) + S_au(t)
\]

where \(u(t) \in \mathbb{R}^m\) represents the control command which is the output of the C&C, \(a(t) \in \mathbb{R}^{m_a}\) denotes a vector describing the effects of unknown cyberattacks on actuators, and \(S_a\) is a matrix of appropriate dimension that indicates the control input channels that are under attack.

The output of the CPS on the C&C side when sensors are under cyberattack can be expressed as

\[
y^s(t) = y_p(t) + D_au(t)
\]

where \(y^s(t) \in \mathbb{R}^p\) denotes the output, \(a(t) \in \mathbb{R}^{p_a}\) denotes the attack signal, and the matrix \(D_a\) describes the sensor attack signature. A CPS in the presence of both the actuator and sensor cyberattacks is depicted in Fig. 1.

Equations (1) and (2) provide a state space realization of the CPS from the C&C side in the following form:

\[
\dot{x}_p(t) = A^s x_p(t) + B^s u^s(t) + B^s_a a(t) + L_1 f_1(t) + N^s \omega^s(t)
\]

\[
y^s(t) = y_p(t) + D_a a(t)
\]

where \(B^s_a = B^s S_a\) is to be interpreted as the actuator cyberattack signature.

In (2) and (3), \(a(t)\) and \(a(t)\) denote the impacts of the adversary’s attack on the control input and output of the CPS, respectively. The signals \(a(t)\) and \(a(t)\) can be arbitrarily changed by the malicious adversary. In the presence of \(a(t)\) and \(a(t)\), the adversary intends to inflict maximum possible damage on the components of the system while simultaneously remaining undetected. The following definitions are needed in the remainder of the article.

**Definition 1 (Weakly Unobservable Subspace [30]):** Let us denote the CPS by \(\Sigma = (A^s, B^s, B^s_a, L_1, N^s, C^s, L_2, D_a)\). Under the fault-free scenario \(f_1(t) = 0\) and \(f_2^s(t) = 0\), the noise-free scenario \(\omega^s(t) = 0\) and \(v^s(t) = 0\), and the cyberattack-free scenario \(\omega^s(t) = 0\) and \(v^s(t) = 0\), a point \(x^s(0) = x_0^s \in \mathbb{R}^n\) is called weakly unobservable if there exists an input function \(u(t)\) such that the output satisfies \(y^s(t) = 0\), \(\forall t \geq 0\). The set of all weakly unobservable points

![Fig. 1. CPS under deception attack on both the input and output channels, where \(u(t)\) denotes the control command, \(a(t)\) represents the cyberattack signal on the input channel, \(v^s(t)\) represents the control input of the plant, \(y_p(t)\) denotes the output on the plant side, \(a(t)\) denotes the attack signal on the output channel, and \(y^s(t)\) denotes the output on the C&C side.](image-url)
is called weakly unobservable subspace and is denoted by \( \mathcal{Y}(\Sigma) \). Moreover, the largest weakly unobservable subspace is denoted by \( \mathcal{Y}^*(\Sigma) \).

Let us denote \( X^i(x^i(0), u(t), a_d(t), a_f(t)) \) as the solution to (4) under the fault-free condition, and \( Y^i(x^i(0), u(t), a_d(t), a_f(t)) = C^iX^i(x^i(0), u(t), a_d(t), a_f(t)) \) as the corresponding output of the CPS, \( \forall t \geq 0 \).

**Definition 2 (Undetectable Cyberattacks [8]):** Given initial conditions \( x_0^i \in \mathbb{R}^n \) and \( x_0^j \in \mathbb{R}^n \), in the CPS (4) under the fault-free scenario, the cyberattack on actuators and sensors using \( u \equiv \{a_d(t), a_f(t)\} \) \( t \geq 0 \), is designated as undetectable if \( Y(x_0^i, u(t), a_d(t), a_f(t)) = Y(x_0^j, u(t), 0, 0), \forall t \geq 0 \), otherwise, the cyberattack is defined as detectable.

**Definition 3 (Input Observable Systems [31]):** The system \( (C, A, B) \) is input observable if \( B \) is monic and \( \text{Im}(B) \) does not intersect the unobservable subspace of \( (C, A) \), where \( \text{Im}(B) \) denotes the image of \( B \).

In the same manner as described in [12] and [31], the sensor fault and sensor noise can be represented by pseudo actuator fault and pseudo process noise, respectively. It is worth noting that in this representation, as described above, sensor faults are mapped into and represented by pseudo actuator faults.

Toward the above end, the following auxiliary invertible LTI system that is driven by the appropriate \( f_2(t) \), which represents the pseudo actuator fault, and \( \omega^a(t) \), which captures the pseudo process noise, is expressed as:

\[
\dot{x}^a(t) = A^a x^a(t) + L^2 f^a_2(t) + N^a \omega^a(t)
\]

\[
C^a x^a(t) = L^2 f^a_2(t) + \nu^a(t)
\]

where \( x^a(t) \in \mathbb{R}^{p_{u}+p_{a}} \), \( f^a_2(t) \in \mathbb{R}^{p_{a}} \), and \( \omega^a(t) \in \mathbb{R}^{p_{a}} \). By incorporating the dynamics of (4) and (5), one can obtain the augmented and extended CPS in the following form:

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Bu(t) + Ba_d(t) + F_1 f_1(t) + F_2 f_2(t) + Na(t) \\
y^a(t) &= Cx(t) + D_a a_d(t)
\end{align*}
\]

where \( x(t) = [x^i(t)^T, x^a(t)^T]^T \), \( A = \text{diag}(A^a, A^s) \), \( B = [B^a_1, 0_{m_x \times (p_{u}+p_{a})}]^T \), \( B_a = [B^a_2, 0_{m_a}]^T \), \( F_1 = [L^1 a(t), 0_{m_a \times (p_{u}+p_{a})}]^T \), \( F_2 = [0_{m_a}, L^a_2]^T \), \( N = \text{diag}(N^a, N^s) \), \( \omega(t) = [\omega^a(t)^T, \omega^s(t)^T]^T \), and \( C = [C^a, C^s] \).

It should be noted that the defined output \( y^a(t) \) in (3) is equal to the one that is given by (6); however, the representations are different.

**B. Model of the Interconnected CPS**

In this section, our objective is to provide a representation of a class of interconnected CPS that are distributed in nature and consist of several subsystems as depicted in Fig. 2. Consequently, considering the large scale of interconnected CPS, one needs to develop a scalable distributed CAFDI methodology that is more desirable for this type of CPS.

We consider the interconnected CPS as consisting of \( N \) subsystems. The dynamics of the \( i \)-th subsystem is expressed as

\[
\begin{align*}
\dot{x}_i(t) &= A_i x_i(t) + \sum_{j \in \mathcal{N}_i} A_{ij} x_j(t) + B_i u^i(t) \\
&\quad + F^i_1 f^i_1(t) + F^i_2 f^i_2(t) + N_i \omega_i(t) \\
y^i_1(t) &= C_i x_i(t) + D_i a_d^i(t), \quad i = 1, \ldots, N
\end{align*}
\]

where \( x_i(t) \in \mathbb{R}^{n_i+p_{u}+p_{a}} \) denotes the state of the \( i \)-th subsystem, \( u^i(t) \in \mathbb{R}^{m_{u}} \) denotes the control input of the subsystem \( i \), \( y^i_1(t) \in \mathbb{R}^{p_{u}} \) denotes the measured output on the C&C side of \( S_i \), \( a_d^i(t) \in \mathbb{R}^{m_{a}} \) denotes the sensor attack signal in the \( i \)-th subsystem, \( \omega_i(t) \in \mathbb{R}^{n_i+p_{u}+p_{a}} \) denotes the mean WSS random Gaussian noise of \( S_i \) with the covariance matrix \( Q_i \), \( f^i_1(t) \in \mathbb{R}^{p_{u}} \) and \( f^i_2(t) \in \mathbb{R}^{p_{a}} \) correspond to actuator and pseudo sensor faults in the subsystem \( i \), respectively, that are derived using a similar method as shown in Section II-A.

Moreover, the matrix \( A_{ij} \) represents the physical coupling between subsystems \( i \) and \( j \in \mathcal{N}_i \), where \( \mathcal{N}_i \) is the set of neighboring subsystems that are coupled with the \( i \)-th subsystem. Furthermore, one has

\[
B_i u^i(t) = B_i u_i(t) + B^i_1 a_d^i(t)
\]

where \( u_i(t) \in \mathbb{R}^{m_{u}} \) denotes the control input generated on the C&C side of \( S_i \), \( a_d^i(t) \in \mathbb{R}^{m_{a}} \) denotes the actuator attack signal in the subsystem \( i \). \( B^i_2 = B^i_2 a^i \) describes the actuator attack signature of the \( i \)-th subsystem, and the matrix \( B^i_1 \) of appropriate dimension indicates the control input channels of \( S_i \) that are compromised by adversaries. The quadruple \((A_i, A_{ij}, B_i, C_i)\) has appropriate dimensions and describes the characteristics of the \( i \)-th subsystem, \( \mathcal{N}_i \) is the noise signature of \( S_i \), and the known pairs \((F^i_1, F^i_2)\) and \((B^i_1, D^i_2)\) capture fault and attack signatures of \( S_i \), respectively.

We consider the following assumptions throughout this article.

**Assumption 1:** In the CPS (1), the number of states is greater than the number of actuators and sensors, i.e., \( n > m \) and \( n > p \).

**Assumption 2:** The adversary has full knowledge on the parameters of (1) and (7), and has access to all the input and output communication channels, i.e., \( m_{a} = m, p_{a} = p, m_{ai} = m_{i}, \) and \( p_{ai} = p_{i} \).

**Remark 1:** It should be emphasized that there may exist local controllers on the plant side of the CPS (6) and the interconnected CPS (7). Hence, the C&C as shown in Figs. 1 and 2 act as outer control loops of the CPS.

**Remark 2:** It should be noted that in the interconnected CPS (7), communication channels among the C&C centers of subsystems can be compromised by adversaries. However, detection of this type of cyberattack is not within the scope of this article and is not addressed here. Methodologies for detecting cyberattacks on the communication channels among the C&C centers of subsystems can be found in [32] and [33].
that are designated to them. We also do not occurrence of anomalies only affects those residual signals is decoupled from all the other anomalies.

Two filters having the same characteristics on both sides are designed in Sections III-A and III-B. Using communication channels, states of the C&C side filter are transmitted to the plant side to generate a residual signal that is sensitive to only cyberattacks while this communication channel may still be compromised by an adversary. However, we assume that there exists a certain number of secure communication channels to transmit states of the C&C side filter to the plant side.

A detector on the plant side that utilizes an UIO is designed in Section III-C. The detector utilizes the previously generated residuals as additional input so that the UIO-based detector is sensitive to both cyberattacks and faults. The reason for selecting UIO as the main detector is that it enables one to utilize a general design structure to simultaneously address the considered CAFDI problems.

Our proposed centralized CAFDI methodology is presented in Section III-D. It is worth noting that by utilizing the proposed methodology, one is still capable of detecting several types of stealthy cyberattacks on the system, such as covert attacks and zero dynamics attacks. Moreover, in Section III-E, the case where all the communication channels between the C&C side filter and the plant side module are nonsecure is studied. Consequently, the C&C side and the plant filters as well as the UIO-based detector are modified to address the CAFDI problem for the CPS in which all the communication channels among the two side filters and detectors are compromised by adversaries.

Our main objective in this article is to address the simultaneous CAFDI problem for the CPS both corresponding to centralized and distributed architectures. Toward this end, we design a bank of observers such that each set of residual signals corresponding to observers is sensitive and specified to detect one specific type of anomaly, namely either an actuator cyberattack \(a_\ell(t)\), a sensor cyberattack \(a_p(t)\), an actuator fault \(f_1(t)\), and/or a pseudo actuator fault \(f_2(t)\), while each residual is decoupled from all the other anomalies.

Decoupling the residuals from one another implies that the occurrence of anomalies only affects those residual signals that are designated to them. We also do not limit our focus on detecting only detectable cyberattacks (see Definition 2).

Our goal and objective is to further detect the so-called undetectable cyberattacks in the sense of Definition 2, e.g., covert and zero dynamics attacks.

First, we assume that the adversary cannot compromise all the communication channels among the proposed C&C side filter and the UIO-based detector, although they have a complete knowledge of parameters of the filters and detectors. Next, we investigate conditions under which adversaries have access to all the communication channels among the C&C side filter and the plant side module are nonsecure.

### C. Objectives

We limit our focus on detecting only detectable cyberattacks (see Definition 2). Our goal and objective is to further detect the so-called undetectable cyberattacks in the sense of Definition 2, e.g., covert and zero dynamics attacks.

### III. CENTRALIZED CAFDI METHODOLOGY

The presence of network layer in the CPS has enabled malicious adversaries to perform cyberattacks on the entire system. On the other hand, due to the existence of this network layer, it is possible to observe the CPS from both the plant side and its C&C side. The idea of observing the CPS from both the plant side and the C&C side is illustrated in Fig. 3. Our goal in this framework is to utilize information from the designed filters on both sides via a communication channel and generate residuals that are specifically sensitive to faults and cyberattacks. Using these residuals, the isolation between faults or cyberattacks can also be achieved.

### A. Command and Control Side Filter

From the C&C side and according to (6), the output of the CPS is governed by

\[
y^*(t) = Cx(t) + Da_s a_s(t). \tag{8}
\]

We have the following standing assumption to be considered throughout this article. **Assumption 3:** Only the communication channels can be compromised and attacked. Consequently, on the C&C side one has access to the control signal, \(u(t)\), before its manipulation by the adversary. Moreover, on the plant side one has access to \(y_p(t)\) before its manipulation by the malicious attacker.

The proposed filter on the C&C side can be expressed as

\[
z^*_c(t) = F^*_p z^*_c(t) + T^*_p Bu(t) + K^*_p y^*(t) \tag{9}
\]

where \(z^*_c(t) \in \mathbb{R}^n\) represents the filter state that estimates \(x^*(t)\) from the C&C side, and the matrices \(F^*_p\), \(T^*_p\), and \(K^*_p\) are of appropriate dimensions that are designed and selected subsequently. The index \(\ell \in \{SA, AA, SF, AF\}\) designates if the filter is designed for detecting sensor attacks, actuator attacks, sensor faults, and actuator faults, respectively.

### B. Plant Side Filter

On the plant side, sensor measurements are carried out before sensor cyberattacks, and the output of the CPS can be expressed as follows:

\[
y_p(t) = Cx(t).
\]

Moreover, on this side one has access to the potentially manipulated control signal \(u^*(t) = u(t) + S_d a_d(t)\).
The proposed filter on the plant side is expressed in the following form:

$$\dot{z}_p^\ell(t) = F_p^\ell z_c^\ell(t) + T_p^\ell Bu^\ell(t) + K_p^\ell y_p(t)$$  \hspace{1cm} (10)

where $z_c^\ell(t) \in \mathbb{R}^n$ denotes the filter state estimating $x_c(t)$ from the plant side. Similar to the C&C side filters, the index $\ell \in \{\text{SA, AA, SF, AF}\}$, indicates if the filter is designed for detecting sensor attacks, actuator attacks, sensor faults, and actuator faults, respectively.

The error signals between estimated states for both sides can be defined as $e_p^\ell(t) = z_c^\ell(t) - z_c^\ell(t)$. The representation of the error dynamics between the two filter states can be derived as follows:

$$\dot{e}_p^\ell(t) = F_p^\ell e_p^\ell(t) + T_p^\ell B_d a_d(t) - K_p^\ell D_d a_d(t).$$  \hspace{1cm} (11)

It follows from (11) that the error dynamics is only sensitive to cyberattacks.

### C. UIO-Based Detector and Residual Signal Generation

We have adopted the UIO design from [13] to develop a UIO-based detector on the plant side with the following representation:

$$\dot{z}_p^\ell(t) = F_p^\ell z_c^\ell(t) + T_p^\ell Bu^\ell(t) + K_p^\ell y_p(t) + L_p^\ell (z_p^\ell(t) - z_c^\ell(t) + D_p q_c(t))$$

$$\dot{x}_p^\ell(t) = z(t)^T + H_p^\ell y_p(t)$$  \hspace{1cm} (12)

where $z_c^\ell(t) \in \mathbb{R}^{n+p+1}$ and $\dot{x}_p^\ell(t) \in \mathbb{R}^{n+p+1}$ denotes the estimated states by the detector, and $q_c(t) \in \mathbb{R}^n$ denotes the cyberattack on the communication channel between the two filters with the signature $D_p$. The matrices $F_p^\ell$, $T_p^\ell$, $K_p^\ell$, $L_p^\ell$, and $H_p^\ell$ are of appropriate dimensions and will be specified subsequently, with $\ell \in \{\text{SA, AA, SF, AF}\}$, denoting the categories defined previously.

The error between the states of the detector and the CPS is defined as $e_p^\ell(t) = x(t) - \dot{x}_p^\ell(t)$. Let

$$\text{res}_p^\ell(t) = y_p(t) - C e_p^\ell(t) \hspace{1cm} (13)$$

denote a residual signal. By selecting $K_1^\ell = K_2^\ell$, $F^\ell = A - H^\ell C A - K_1^\ell C$, with $K_1^\ell$ of appropriate dimension, and $K_2^\ell = F^\ell H^\ell$, the dynamics associated with $e_p^\ell(t)$ can now be expressed in the following form:

$$\dot{e}_p^\ell(t) = (A - H^\ell C A - K_1^\ell C) e_p^\ell(t) + (I - T_p^\ell - H_p^\ell C) \times (Bu(t) + B_d a_d(t)) + (I - H_p^\ell C) F_1 f_1(t) + (I - H_p^\ell C) F_2 f_2(t) + (I - H_p^\ell C) N_\omega(t) - L_p^\ell e_p^\ell(t) - L_p^\ell D_p q_c(t).$$  \hspace{1cm} (14)

**Definition 4:** A cyberattack/fault is detected if the residual signal $\text{res}_p^\ell(t)$ given by (13) exceeds a prespecified threshold $\eta > 0$ as follows:

$$\|\text{res}_p^\ell(t)\|_2 > \eta$$

where $\|\cdot\|_2$ indicates the Euclidean norm.

**Remark 3:** To select the threshold $\eta$, one may need to perform Monte Carlo simulation runs for the healthy system, i.e., for the fault-free and cyberattack-free system in presence of external disturbances and noise and choose the maximum value of $\|\text{res}(t)\|_2$ as $\eta$.

**Definition 5 (Decoupled Residual):** The residual signal $\text{res}_p^\ell(t)$ given by (13) is decoupled from an anomalous signal in the set $\{a_d(t), a_c(t), a_{cp}(t), f_1(t), f_2(t)\}$ if the dynamics and trajectory of $\text{res}_p^\ell(t)$ are not affected by that anomalous signal.

The following assumption stands throughout this section.

**Assumption 4:** The malicious adversary knows the parameters of the C&C side filter in (9), the plant side filter in (10), and the UIO-based detector in (12).

### D. Filters and Detectors Design for CAFDI Objectives

The error dynamics in (11) and (14) can now be augmented as follows:

$$\dot{\tilde{e}}^\ell(t) = \tilde{F}^\ell \tilde{e}^\ell(t) + \tilde{B}^\ell u(t) + \tilde{B}^\ell_0 a_d(t) + \tilde{F}^\ell_1 f_1(t) + \tilde{F}^\ell_2 f_2(t) - K_p^\ell_d a_d(t) - L_p^\ell a_{cp}(t) + \tilde{N}^\ell \omega(t)$$  \hspace{1cm} (15)

where $\tilde{e}^\ell(t) = [e_p^\ell(t)^T]^T$, and

$$\tilde{F}^\ell = \begin{bmatrix} F_p^\ell & -L_p^\ell \\ 0 & 0 \end{bmatrix}, \quad \tilde{B}^\ell = \begin{bmatrix} (I - T_p^\ell - H_p^\ell C) B \\ 0 \end{bmatrix}$$

$$\tilde{B}^\ell_0 = \begin{bmatrix} (I - T_p^\ell - H_p^\ell C) B_a \\ T_p^\ell B_a \end{bmatrix}, \quad \tilde{F}^\ell_1 = \begin{bmatrix} (I - H_p^\ell C) F_1 \\ 0 \end{bmatrix}$$

$$\tilde{F}^\ell_2 = \begin{bmatrix} (I - H_p^\ell C) F_2 \\ 0 \end{bmatrix}, \quad \tilde{K}_p^\ell = \begin{bmatrix} K_p^\ell D_a \\ 0 \end{bmatrix}, \quad \tilde{L}^\ell = \begin{bmatrix} L_p^\ell D_p \\ 0 \end{bmatrix}$$

$$\tilde{N}^\ell = \begin{bmatrix} 0 \\ (I - H_p^\ell C) N \end{bmatrix}$$  \hspace{1cm} (16)

where $\ell \in \{\text{SA, AA, SF, AF}\}$.

**Assumption 5:** There exist $q = \max\{m_a, p_a\}$ secure communication channels among the C&C side filter in (9) and the UIO-based detector in (12), i.e., $rank(D_p) = n - q$. Moreover, $C_q = \{c_1, \ldots, c_q\}$ denotes the set of secured communication channels, where $c_q \in \{1, \ldots, n\}$, for $q \in \{1, \ldots, q\}$.

In the following, it is shown how one can generate four residual signals $\text{res}_{AA}(t)$, $\text{res}_{SA}(t)$, $\text{res}_{AF}(t)$, and $\text{res}_{SA}(t)$ to detect the actuator cyberattack, the sensor cyberattack, the actuator fault, and the sensor fault, respectively, using a bank of filters and four UIO-based detectors.

**Proposition 1:** Under Assumption 5, the residual signal $\text{res}_{AA}(t) = y_p(t) - C_1 e_p(t)$ is affected by the actuator cyberattacks $a_d(t)$ and is decoupled from $a_d(t)$, $a_{cp}(t)$, $f_1(t)$, and $f_2(t)$ in the sense of Definition 5, if the following conditions for the augmented dynamics (15) hold for $\ell = \text{AA}$, namely:

1) $T_p^\ell = I - H_p^\ell C$;
2) $(I - H_p^\ell C) F_1 = 0$;
3) $(I - H_p^\ell C) F_2 = 0$;
4) $L_p^\ell D_p = 0$;
5) $K_p^{\text{AA}} D_a = 0$;
6) the triplet $(C, F^\ell, \tilde{L}^\ell)$ is left-invertible, where $\tilde{L}^\ell = [L_{c_1}^\ell, \ldots, L_{c_q}^\ell]$, and $L_{c_q}^\ell$ is the $c_q$th column of $L^\ell$, for $q \in \{1, \ldots, q\}$;
7) the Rosenbrock system matrix

$$P_{\Sigma_a}(s) = \begin{bmatrix} sI - F_{\text{AA}}^\ell & -T_p^\text{AA} B_a \\ L_{\text{AA}}^\ell & 0_{(n+p_a+p) \times m_a} \end{bmatrix}$$

does not have any non-minimum phase zero dynamics;
8) $rank(L_{\text{AA}}^\ell T_p^\text{AA} B_a) = rank(T_p^\text{AA} B_a)$;
Equation (23) implies that $\text{Im}(T_p^{AA}B_a)$ should not be in the null space of $L^{AA}$, which is equivalent to
\[
\text{rank} \left( L^{AA}T_p^{AA}B_a \right) = \text{rank} \left( T_p^{AA}B_a \right).
\] (24)

Given that as per Assumption 5 one has $\text{rank}(L^{AA}) \geq \text{rank}(T_p^{AA}B_a)$, therefore the matrix $L^{AA}$ can be obtained such that (24) holds.

The Rosenbrock system matrix $P_{\Sigma}(s)$ being left-invertible implies that for any $a_t(0) \neq 0$, $L^{AA}P_{\Sigma}(s) \neq 0$.

Finally, to detect actuator cyberattacks, the governing dynamics in (17) should be stable. This completes the proof of the Proposition 1.

**Proposition 2:** Under Assumption 5, the residual signal $\text{res}_{SA}(t) = y_t(s) - C \hat{x}^{SA}(t)$ is affected by the sensor cyberattacks $a_t(s)$ and is decoupled from $a_t(s)$, $a_{cp}(t)$, $f_1(t)$, and $f_2(t)$ in the sense of Definition 5, if Conditions 1)–4), 6), and 9) of Proposition 1 for $\ell = SA$, and the following conditions for the augmented error dynamics (15) hold, namely:

1) $T_p^{SA}B_a = 0$;
2) the Rosenbrock system matrix
\[
P_{\Sigma(a)}(s) = \begin{bmatrix} sI - F_p^{SA} & K_p^{SA}D_a \\ L^{SA} & 0_{(n+\ell+p+q)\times(p+q)} \end{bmatrix}
\]
does not have any nonminimum phase zero dynamics;
3) rank$(L^{AA}K_p^{SA}D_a)$ = rank$(K_p^{SA}D_a)$.

**Proof:** The proof follows along similar lines to that of Proposition 1 and is omitted for sake of brevity.

**Remark 4:** Suppose Condition 8) of Proposition 1 is not satisfied and $P_{\Sigma}(s)$ is not left-invertible. In this case, it has been shown in [7] that one can find an actuator cyberattack $a_t(s) \neq 0$ such that $L^{AA}e^{AA}(t) = 0$. This type of cyberattack has been represented in [7] and has been defined as “undetectable controllable attacks” in [8]. According to (18) and (19) the actuator cyberattack signal $a_t(s)$ can affect the error $e^{AA}(t)$ only through $L^{AA}e^{AA}(t)$. Hence, the adversary has the capability of injecting a stealthy cyberattack using $a_t(s)$ that does not affect the residual signal $\text{res}_{SA}(t) = C \hat{x}^{AA}(t)$. Similarly, it can be shown that if Condition 3) of Proposition 2 is not satisfied and $P_{\Sigma(s)}$ is not left-invertible, the adversary can inject stealthy attack using $a_t(s)$ which does not affect the residual $\text{res}_{SA}(t)$.

**Remark 5:** In Propositions 1 and 2, there is no assumption on the nature, characteristics, and type of sensor and actuator cyberattacks. This implies that using the proposed methodology, one is capable of detecting and isolating detectable attacks, such as false data injection attacks, as well as undetectable attacks (refer to Definition 2), such as covert attacks and zero dynamics attacks. Furthermore, since as per Definition 4, we are using a threshold checking mechanism to make a decision on the anomalous status of the CPS, it would be still possible for adversaries to design their attack signals such that the residual remains below the threshold. Hence, in such a scenario, adversaries will try to reduce the amplitude of their attack signals to remain undetected which implies that the attack signals may not necessarily lead the CPS to dangerous conditions.

**Proposition 3:** The residual signal $\text{res}_{SA}(t) = y_t(s) - C \hat{x}^{AF}(t)$ is affected by the actuator fault $f_1(t)$ and is decoupled from $a_t(s)$, $a_{cp}(t)$, and $f_2(t)$ in the sense of
Definition 5, provided that $L^{AF} = 0$ and Conditions 1), 3), and 9) of Proposition 1 hold for $\ell = AF$.

Proof: In light of the Conditions 1) and 3) of Proposition 1, and setting $\ell = AF$, (15) yields

$$\hat{e}^{AF}(t) = \tilde{e}^{AF} + \tilde{f}_a f_1(t) - \tilde{K}_p a(t) - \tilde{L}^{AF} a(t) + \tilde{N}^{AF} \omega(t).$$

Moreover, by setting $L^{AF} = 0$, the dynamics of $e^{AF}(t)$ is governed by

$$e(t) = F^{AF} e^{AF}(t) + (I - H^{AF} C) F_1 f_1(t) + N(t) \omega(t)$$

and consequently, the residual signal $\text{res}^{AF}(t) = C^{AF} e(t)$ is only sensitive to the actuator fault $f_1(t)$. In addition, $F^{AF}$ should be Hurwitz to have a stable error dynamics $e^{AF}(t)$. This completes the proof of the Proposition 3.

**Proposition 4:** The residual signal $\text{res}^{SF}(t) = \gamma_p(t) - C^{SF} f_2(t)$ is affected by the pseudo actuator fault $f_2(t)$ and is decoupled from $a(t)$, $a(t)$, $\omega(t)$, $f_1(t)$ in the sense of Definition 5, provided that $L^{SF} = 0$ and Conditions 1), 2), and 9) of Proposition 1 hold for $\ell = SF$.

Proof: Setting $\ell = SF$, the proof follows along similar lines to that of Proposition 3 and is omitted for sake of brevity.

As stated in [13], the Conditions 2) and 3) in the Proposition 1 are solvable if and only if rank($C F_1$) = rank($F_1$); and rank($C F_2$) = rank($F_2$). The next theorem provides sufficient conditions for isolability of sensor and actuator faults.

**Theorem 1:** The residuals $\text{res}^{AF}(t)$ and $\text{res}^{SF}(t)$ can be simultaneously generated to detect and isolate $f_1(t)$ and $f_2(t)$ if $F_1^T F_2 = 0$.

Proof: To generate the residual signal $\text{res}^{AF}(t)$, the Condition 2) in Proposition 3 should hold, which can be interpreted as requiring

$$\text{Im}(I - H^{AF} C) \subset \ker(F_1^T)$$

and at the same time, the impact of $f_1(t)$ should show up in the dynamics of $e(t)$, that is, $\text{Im}(I - H^{AF} C) F_1 \neq 0$. The latter condition is equivalent to having

$$\text{Im}(F_1^T) \subset \ker(I - H^{AF} C).$$

From (25) to (26), it can be inferred that $\text{Im}(F_1^T) \subset \ker(F_2^T)$, which implies that $F_1^T F_2 = 0$. Note that the case of generating the residual signal $\text{res}^{SF}(t)$ provides one with the same result. This completes the proof of Theorem 1.

It follows from the definitions of $F_1$ and $F_2$ that the condition $F_1^T F_2 = 0$ is always satisfied. Therefore, as long as Conditions 2) and 3) in Proposition 1 are solvable, the actuator faults and pseudo actuator faults can be detected and isolated.

**Remark 6:** Given that $L^{AF}$ and $L^{SF}$ are equal to zero in Propositions 3 and 4, to generate the residual signals $\text{res}^{SA}(t)$, $\text{res}^{SF}(t)$, and $\text{res}^{SF}(t)$ one needs to construct a bank of four filters (two on each side) with the states $z_p^{AF}(t)$, $z_c^{AF}(t)$, and $z_p^{SF}(t)$, and $z_c^{SF}(t)$, and four UIO-based detectors with the states $\tilde{x}_p^{AF}(t)$, $\tilde{x}_c^{AF}(t)$, $\tilde{x}_p^{SF}(t)$, and $\tilde{x}_c^{SF}(t)$ according to Propositions 1–4. In Propositions 1 and 2, the matrices $K_p^{SA}$ and $T^{SA}$ have been utilized to decouple sensor cyberattacks and actuator cyberattacks in the sense of Definition 5 from the generated residual signals, respectively. Hence, one can conclude that there is no contradiction among the conditions to generate $\text{res}^{SA}(t)$ and $\text{res}^{SA}(t)$. Subsequently, from Theorem 1 it can be seen that no contradiction exists among the design conditions in Propositions 3 and 4 to generate $\text{res}^{AF}(t)$ and $\text{res}^{SF}(t)$. Moreover, in Propositions 3 and 4, the matrix $L^\ell$ has been employed to decouple the cyberattacks from $\text{res}^{AF}(t)$ and $\text{res}^{SF}(t)$, which indicates that there are no contradictions in the design conditions for Propositions 1–4.

**Remark 7:** One application of the proposed centralized CAFDI methodology could be the detection and isolation of anomalies in a single UAV. Consider a UAV that is remotely controlled and receives its way points from the C&C center. An adversary is capable of performing man-in-the-middle cyberattack to either hijack or destroy the UAV. Considering the availability of relatively cheap and powerful microcontrollers, it is reasonable to assume that a UAV can have adequate computational resources to deploy and run the proposed UIO-based detector and filters in its plant side. Moreover, given that in our proposed centralized CAFDI methodology, detection of anomalies occurs in the plant side, this information can be relayed back to the C&C side as a flag for corrective actions to be considered.

**E. The Case of Fully Nonsecure Communication Channels**

In Section III-D under Assumption 5, we considered the existence of secure communication channels between the two side filters. Consequently, the generated residuals in Propositions 1–4 were decoupled from the communication channel attack signal $a(t)$. However, it is possible that adversaries compromise all the communication channels among the two side filters. Hence, in this section we remove Assumption 5 and consider the case where there exists no secure communication channel between the filters. Furthermore, the proposed filters in (9) and (10) are modified to address the CAFDI problem.

To develop the modified filters, one requires two communication channels, one from the C&C side filter to the plant side filter and the other from the plant side filter to the C&C side filter to transmit states of the two filters to one another. Moreover, the specified communication channels are assumed to be fully compromised by the adversaries.

The proposed filter on the plant side is modified in the following form:

$$\dot{z}_p^\ell(t) = F_p^\ell z_p^\ell(t) + T_p^\ell Bu^\ast(t) + K_p^\ell y_p(t) + \delta_p(t)L_p^\ell \left(\tilde{z}_p^\ell(t) - \tilde{z}_c^\ell(t) + D_p^\ell \delta_p(t)\right)$$

where $\delta_p(t)$ denotes a scalar random variable in the interval $[\delta_1, \delta_2]$, $\delta_1$ and $\delta_2$ are positive real numbers, $z_p^\ell(t)$ denotes the state of the C&C side filter transmitted to the plant side, and $a(t) \in \mathbb{R}^{nc}$ denotes the cyberattack on the communication channel from the C&C side filter to the plant side filter with the signature $D_p$. Consequently, the modified filter on the C&C side can be expressed as

$$\dot{z}_c^\ell(t) = F_c^\ell z_c^\ell(t) + T_p^\ell Bu^\ast(t) + K_p^\ell y_p(t) + \delta_c(t)L_p^\ell \left(\tilde{z}_c^\ell(t) - \tilde{z}_c^\ell(t) + D_p^\ell \delta_p(t)\right)$$

where $\delta_c(t) \in [\delta_1, \delta_2]$ denotes a scalar random variable and $a(t) \in \mathbb{R}^{nc}$ denotes the cyberattack on the communication channel.
channel from the plant side filter to the C&C side filter associated with the signature $D_{pc}$.

Assumption 6: The adversary has access to all the communication channels among the two side filters, i.e., $D_{cp} = D_{pc} = I_{n}$.

Considering Assumption 6, and the dynamics (27) and (28), the governing dynamics of the error $e_{\ell}^{p}(t) = z_{\ell}^{p}(t) - z_{\ell}^{t}(t)$ can be derived as follows:

$$
\dot{e}_{\ell}^{p}(t) = F_{p}^{\ell}(t)e_{p}^{\ell}(t) + T_{p}^{\ell}B_{a}a_{t}(t) - K_{p}^{\ell}D_{a}a_{t}(t) - \delta_{p}(t)L_{p}^{\ell}a_{cp}(t) + \delta_{c}(t)L_{p}^{\ell}a_{pc}(t)
$$

(29)

where $F_{p}^{\ell}(t) = F_{p}^{\ell} + \delta(t)L_{p}^{\ell}$, and $\delta(t) = \delta_{p}(t) + \delta_{c}(t)$.

Moreover, the dynamics of the UIO-based detector on the plant side, as given in (12), can now be rewritten in the following form:

$$
\dot{z}_{\ell}^{t}(t) = F_{t}^{\ell}z_{\ell}^{t}(t) + T_{t}^{\ell}B_{u}u(t) + K_{t}^{\ell}y_{p}(t)
$$

$$
+ L^{\ell}(e_{p}^{\ell}(t) - (z_{\ell}^{p}(t) + a_{cp}(t)))
$$

(30)

Consequently, the error dynamics of $e^{\ell}(t) = x(t) - \dot{x}(t)$ in (14) can be reformulated as follows:

$$
\dot{e}_{\ell}^{t}(t) = (A - H^{t}CA - K_{1}^{t}C)e_{\ell}^{t}(t) + (I - T^{t} - H^{t}C)
$$

$$
\times \left( Bu(t) + B_{a}a_{t}(t) + \left( I - H^{t}C \right) F_{1}f_{1}(t) + \left( I - H^{t}C \right) F_{2}f_{2}(t)
$$

$$
+ \left( I - H^{t}C \right) N_{o} \omega(t) - L^{\ell}(e_{p}^{\ell}(t) - a_{cp}(t)) \right)
$$

(31)

Assumption 7: The malicious adversary does not have knowledge on the parameters $\delta_{p}(t)L_{p}^{\ell}$ in (27) and $\delta_{c}(t)L_{p}^{\ell}$ in (28), however, the remaining parameters in (27), (28), and (30) are known to the adversary.

Remark 8: Given the randomness of the variables $\delta_{p}(t)$ and $\delta_{c}(t)$, it is reasonable to assume that adversaries do not know the values of the parameters stated in Assumption 7. Moreover, $\delta_{p}(t)$ and $\delta_{c}(t)$ can have any arbitrary probability distributions associated with them.

Lemma 1: Let Assumptions 6 and 7 hold and consider the CPS (6) under cyberattacks and faults. Given the modified plant side filter (27) and the modified C&C side filter (28), adversaries cannot design communication channel attack signals $a_{cp}(t)$ and $a_{pc}(t)$ to ensure $L^{\ell}(e_{p}^{\ell}(t) - a_{cp}(t)) = 0$, $\forall t > 0$ in the error dynamics (31).

Proof: We consider three possible scenarios, namely $a_{cp}(t) = 0$ and $a_{pc}(t) \neq 0$, $a_{pc}(t) = 0$ and $a_{cp}(t) \neq 0$, and finally $a_{cp}(t) \neq 0$ and $a_{pc}(t) \neq 0$. Let $a_{cp}(t) = 0$ and consider the error dynamics $e_{p}^{\ell}(t)$ in (29) with the output $L^{\ell}e_{p}^{\ell}(t)$. Given the unknown random parameter $\delta_{c}(t)L_{p}^{\ell}$ in the dynamics of $e_{p}^{\ell}(t)$, adversaries cannot design $a_{cp}(t) = 0$ such that $-K_{p}^{\ell}D_{a}a_{t}(t) + \delta_{c}(t)L_{p}^{\ell}a_{cp}(t) = 0$ or $T_{p}^{\ell}B_{a}a_{t}(t) + \delta_{c}(t)L_{p}^{\ell}a_{pc}(t) = 0$. Moreover, since adversaries do not know $F_{p}^{\ell}(t)$ and $\delta_{c}(t)L_{p}^{\ell}$, they cannot execute zero dynamics attacks or perform "undetectable controllable attacks" (refer to Remark 4) on the triplet $(L^{\ell}, F_{p}^{\ell}(t), \delta_{c}(t)L_{p}^{\ell})$. A similar argument can be used to show that adversaries cannot make $L^{\ell}(e_{p}^{\ell}(t) - a_{cp}(t)) = 0$, $\forall t > 0$, in the case of $a_{pc}(t) = 0$ and $a_{cp}(t) = 0$.

Let $a_{cp}(t) = 0$ and $a_{pc}(t) = 0$. Since the parameter $F_{p}^{\ell}(t)$ is unknown to adversaries, they cannot design the communication attack signal $a_{cp}(t)$ to eliminate the impact of cyberattacks in (29) from the signal $L^{\ell}(e_{p}^{\ell}(t) - a_{cp}(t))$. This completes the proof of the lemma.

Proposition 5: Let Assumptions 6 and 7 hold and consider the modified plant side filter (27), the modified C&C side filter (28), and the UIO-based detector (30), where $\ell = AA$. The residual signal $r_{AA}(t) = y_{p}(t) - C^{AA}a_{t}(t)$ is affected by $a_{a}(t)$, $a_{cp}(t)$, and $a_{pc}(t)$ and is decoupled from $a_{q}(t)$, $f_{1}(t)$, and $f_{2}(t)$ in the sense of Definition 5 provided that the Conditions 1–3), and 5) of Proposition 1 and the following conditions hold for the error dynamics (29) (31), namely:

1. the triplet $(C, F^{\ell}, L^{\ell})$ is input observable in the sense of Definition 3;
2. $F^{\ell}$ is Hurwitz;
3. $F_{p}^{\ell}(t)$ is designed such that there exists a symmetric positive definite matrix $Q_{p}^{\ell}(t)$ that satisfies

$$
F_{p}^{\ell}(t)^{T} + F_{p}^{\ell}(t) = -Q_{p}^{\ell}(t)
$$

(32)

where $\beta_{cp}I_{n} \leq Q_{p}^{\ell}(t)$, and $\beta_{cp}$ is a positive scalar.

Proof: Let $\ell = AA$. Under the Conditions 1–3), and 5) in Proposition 1, the error dynamics (29) and (31) become

$$
\dot{e}_{\ell}^{AA}(t) = F_{p}^{AA}(t)e_{p}^{AA}(t) + T_{p}^{AA}B_{a}a_{t}(t) - \delta_{p}(t)L_{p}^{AA}a_{cp}(t) + \delta_{c}(t)L_{p}^{AA}a_{pc}(t)
$$

(33)

and

$$
e_{p}^{AA}(t) = F_{p}^{AA}(t)e_{p}^{AA}(t) - L_{p}^{AA}(e_{p}^{AA}(t) - a_{cp}(t)) - N_{o} \omega(t) + (I - H^{AA}C)N_{o} \omega(t)
$$

(34)

respectively.

Consider the error dynamics (33) with the output $L_{p}^{AA}(e_{p}^{AA}(t) - a_{cp}(t))$. Given that adversaries do not know $F_{p}^{AA}(t)$, by utilizing the actuator attack signal $a_{q}(t)$, they cannot excite the zero dynamics of the triplet $(L_{p}^{AA}, F_{p}^{AA}(t), T_{p}^{AA}B_{a})$ or carry out "undetectable controllable attacks" (refer to Remark 4). Moreover, according to Lemma 1, and given that $(C, F^{\ell}, L^{\ell})$ is input observable, the impact of cyberattacks $a_{a}(t)$, $a_{cp}(t)$, and $a_{pc}(t)$ in (33) will be manifested in the residual signal $r_{AA}(t) = Ce_{AA}(t)$ through the error dynamics (34).

Consequently, to detect cyberattacks, one needs to show that the error dynamics (34) and (33) are stable. The dynamics (34) is stable if $F_{p}^{AA}$ is Hurwitz. Furthermore, consider the Lyapunov function candidate $V_{p}(e_{p}^{AA}(t)) = e_{p}^{AA}(t)^{T}e_{p}^{AA}(t)$.

The derivative of $V_{p}(e_{p}^{AA}(t))$ along the trajectories of (33) can be obtained as

$$
\dot{V}_{p}(e_{p}^{AA}(t)) = e_{p}^{AA}(t)^{T}e_{p}^{AA}(t) + e_{p}^{AA}(t)^{T}e_{p}^{AA}(t) - e_{p}^{AA}(t)^{T}Q_{p}^{AA}(t)e_{p}^{AA}(t).
$$

(35)

It follows from (35) that $\dot{V}_{p}(e_{p}^{AA}(t)) \leq -\beta_{cp}\|e_{p}^{AA}(t)\|^{2}$, which implies that under Condition 2), the error dynamics (33) is stable [35]. This completes the proof of the proposition.

Remark 9: According to the Condition 3) in Proposition 5, one needs to design $F_{p}^{\ell}(t)$ such that (32) holds. Given that $F_{p}^{\ell}(t) = F_{p}^{\ell} + \delta(t)L_{p}^{\ell}$, one has

$$
F_{p}^{\ell}(t) + F_{p}^{\ell}(t)^{T} = F_{p}^{\ell} + \delta(t)L_{p}^{\ell}
$$

where $F_{p}^{\ell} = F_{p}^{\ell} + F_{p}^{\ell}$ and $L_{p}^{\ell} = L_{p}^{\ell} + L_{p}^{\ell}$ are symmetric matrices. Thus, one can use $F_{p}^{\ell}$ to design $F_{p}^{\ell}$ such that
$\hat{F}_p + 2\delta_1 \hat{L}_p^\ell$ and $\hat{F}_p^\ell + 2\delta_2 \hat{L}_p^\ell$ are negative definite, which is the sufficient condition for (32) to hold.

**Proposition 6:** Under Assumptions 6 and 7, the residual signal $\text{res}_{SA} = y_p(t) - C^\ell \hat{x}_S(t)$ generated by utilizing the modified plant side filter (27), the modified C&C side filter (28), and the UIO-based detector (30) is affected by $a_{\ell}(t)$, $a_{c\ell}(t)$, and $a_{pc}(t)$ and is decoupled from $a_{\ell}(t)$, $f_1(t)$, and $f_2(t)$ in the sense of Definition 5 provided that the Conditions 1)–3) of Proposition 1, the Condition 1) of Proposition 2, and the Conditions 1)–3) of Proposition 5 for $\ell = SA$ hold.

**Proof:** The proof follows along similar lines to that of Proposition 5 and is omitted for sake of brevity.

**Remark 10:** In the Proposition 1, the residual signal is only sensitive to the actuator attack signal $a_{\ell}(t)$, however, due to not having any secure communication channel in Proposition 5, the residual signal is affected by the set of signals $\{a_{\ell}(t), a_{c\ell}(t), a_{pc}(t)\}$. Similarly, the residuals in Propositions 2 and 6 are affected by the sensor attack signal $a_{\ell}(t)$ and the set $\{a_{\ell}(t), a_{c\ell}(t), a_{pc}(t)\}$. Therefore, there is no secure communication channel, the generated residuals in Propositions 5 and 6 cannot be decoupled from cyberattacks on the communication channels between the two side filters. Moreover, the given conditions in Propositions 3 and 4 can be used to design and implement actuator and sensor FDI modules, respectively. In other words, having nonsecure communication channels among the two side filters does not affect the performance of our proposed FDI methodologies and modules.

IV. DISTRIBUTED CAFDI METHODOLOGY FOR INTERCONNECTED CPS

In this section, our objective is to extend our results in Section III-E and address the CAFDI problem for large scale interconnected CPS given by (7) through a distributed architecture. The proposed CAFDI methodology is distributed in the sense that CAFDI modules on each subsystem communicate information with their neighboring subsystems. Hence, each subsystem can detect and isolate its cyberattacks and faults as well as anomalies in its neighboring subsystems. Furthermore, we consider the detection of both detectable and undetectable cyberattacks (refer to Definition 2) in our proposed methodology.

**A. UIO-Based Detectors and Filters Design for the $i$th Subsystem**

In this section, we use a similar approach as in Section III-E to design both side filters and UIO-based detectors. Each subsystem is equipped with a bank of filters both on its plant side and its C&C side, where both side filters transmit their states to one another over compromised communication channels. In this methodology, we consider the existence of one secure communication channel among the two side filters. Moreover, a detector using the UIO is designed and utilized on the plant side of each subsystem. Each UIO-based detector receives estimated states of the UIO-based detectors in its neighborhood through compromised communication channels. It should be noted that in our proposed distributed CAFDI methodology for interconnected CPS, two side filters of each subsystem do not transmit information to the nearby filters that are in the other subsystems. The proposed distributed CAFDI methodology is depicted in Fig. 4.

Consequently, the proposed filter on the plant side of $S_i$ is given by

$$\dot{z}_{pi}^\ell(t) = F_{pi}^\ell z_{pi}^\ell(t) + T_{pi}^\ell B_{pi} u_{pi}^\ell(t) + K_{pi}^\ell y_{pi}^\ell(t) + \delta_{pi}(t) L_{pi}^\ell \times \left( z_{pi}^\ell(t) - \left( \hat{z}_{pi}^\ell(t) + D_{pc_{pi}^\ell}(t) \right) \right)$$

where $z_{pi}^\ell(t) \in \mathbb{R}^{n_i}$ denotes the state of the plant side filter on the $i$th subsystem, $y_{pi}^\ell(t) = C_{pi}^{\ell} x_{pi}^\ell(t)$ denotes the measured output of $S_i$ on the plant side, $z_{pi}^\ell(t) \in \mathbb{R}^{n_i}$ denotes the state of the C&C side filter of $S_i$ which is transmitted to the plant side, $a_{c\ell}^\ell(t) \in \mathbb{R}^{n_{ci}}$ denotes the cyberattack on the communication channel from the C&C side filter of $S_i$ to its plant side filter with the signature $D_{pc_{ci}^\ell}$ and $\delta_{pi}(t) \in [\delta_1, \delta_2]$ denotes a scalar random variable.

Moreover, the proposed filter on the C&C side of the $i$th subsystem is governed by

$$\dot{z}_{ci}^\ell(t) = F_{ci}^\ell z_{ci}^\ell(t) + T_{ci}^\ell B_{ci} u_{ci}^\ell(t) + K_{ci}^\ell y_{ci}^\ell(t) + \delta_{ci}(t) L_{ci}^\ell \left( z_{ci}^\ell(t) - \left( z_{ci}^\ell(t) + D_{pc_{ci}^\ell}(t) \right) \right)$$

where $\delta_{ci}(t) \in [\delta_1, \delta_2]$ is a scalar random variable, and $a_{pc_{ci}^\ell}^\ell(t) \in \mathbb{R}^{n_{ci}}$ denotes the cyberattack on the communication channel from the plant side filter of $S_i$ to its C&C side filter with the signature $D_{pc_{ci}^\ell}$.

Let us define the error $e_{pi}^\ell(t) = z_{pi}^\ell(t) - z_{ci}^\ell(t)$ for the $i$th subsystem. The dynamics of the error $e_{pi}^\ell(t)$ can be derived in the following form:

$$\dot{e}_{pi}^\ell(t) = F_{pi}^\ell e_{pi}^\ell(t) + T_{pi}^\ell B_{pi} a_{pc_{pi}^\ell}(t) - K_{pi}^\ell D_{pc_{pi}^\ell}(t) \times$$

$$- \delta_{pi}(t) L_{pi}^\ell D_{pc_{pi}^\ell}(t) + \delta_{ci}(t) L_{ci}^\ell D_{pc_{ci}^\ell}(t)$$

where $F_{pi}^\ell(t) = F_{pi}^\ell + \delta_{ci}(t) L_{ci}^\ell + \delta_{pi}(t) L_{pi}^\ell$.

The UIO-based detector of $S_i$ can be expressed in the following form:

$$\dot{z}_{pi}^\ell(t) = F_{pi}^\ell z_{pi}^\ell(t) + T_{pi}^\ell B_{pi} u_{pi}^\ell(t) + K_{pi}^\ell y_{pi}^\ell(t)$$

$$+ \sum_{j \in N_i} A_{ij} \left( \hat{x}_j^\ell(t) + D_{pp_{ij}^\ell}(t) \right)$$

$$+ \delta_{pi}(t) L_{pi}^\ell \left( z_{pi}^\ell(t) - \left( z_{ci}^\ell(t) + D_{pc_{pi}^\ell}(t) \right) \right)$$

where $\delta_{pi}(t) \in [\delta_1, \delta_2]$ denotes a random variable, $\hat{x}_j^\ell(t) = z_{pi}^\ell(t) + H_{ji}^\ell y_{pi}^\ell(t)$ denotes the estimation of the state in the
jth subsystem which is transmitted through a communication channel to the UIO-based detector in $S_i$, and $a_{ij}^{(d)}(t) \in \mathbb{R}^{q_{ij}}$ denotes the malicious attack signal on the communication channel among $S_i$ and $S_j$ with the signature $D_{pp}^i$, for $j \in N_i$. Moreover, the matrices $F_i$, $F_i^{(j)}$, $L_i^e$, and $H_i^e$ are of appropriate dimensions and satisfy $K_i^{(1)} + K_i^{(2)} = A_i - H_i^e C_i A_i - K_i^{(1)} C_i$, where $K_i^{(1)}$ is a matrix of appropriate dimension, and $K_i^{(2)} = F_i^e H_i^e$.

Consequently, the error dynamics of $e_i^e(t) = x_i(t) - \hat{x}_i^e(t)$ in the subsystem $i$ can be expressed by

$$
\dot{e}_i^e(t) = F_i^e e_i^e(t) + (I - T_i^e - H_i^e C_i)(B_i u_i(t) + B_i^u a_{ij}^u(t)) + (I - H_i^e C_i) \sum_{j \in N_i} (A_{ij} (e_j^e(t) - D_{pp}^i a_{ij}^{(d)}(t)))
+ (I - H_i^e C_i)(F_i^e f_i^j(t) + F_i^e f_i^j(t)) - \delta_i(t) L_i^e (e_i^e(t) - D_{cpi} a_{ci}(t))
$$

(40)

where $e_i^e(t) \in \mathbb{R}^{n_i + p + n_i}$, and $\hat{x}_i^e(t) \in \mathbb{R}^{n_i + p_i + p_i}$ is the estimated states by the detector of $S_i$.

Moreover, by utilizing (39), one can generate a residual signal on the plant side of $S_i$ in the following form:

$$
\text{res}_i^a(t) = y_{pi}(t) - C_i \hat{x}_i^e(t) = C_i e_i^e(t).
$$

(41)

**Definition 6:** A cyberattack/fault on the $i$th subsystem is detected if the following inequality is satisfied for the residual signal (41):

$$
\left\| \text{res}_i^a(t) \right\|_2 > \eta_i^a
$$

where $\eta_i^a > 0$ is a pre-specified threshold.

**Similar to the value of $\eta$** in Definition 4 (see also Remark 3), the prescribed threshold $\eta_i^a$ can be computed by means of Monte Carlo simulation runs for the healthy system.

**Definition 7:** The generated residual signal in (41) which belongs to the $i$th subsystem is decoupled from an anomalous signal in the set $\{a_{ij}^u(t), a_{ij}^{(d)}(t), a_{ip}^u(t), a_{ip}^{(d)}(t), f_i^j(t), f_i^j(t)\}$ if that anomalous signal does not affect the dynamics and trajectories of $\text{res}_i^a(t)$.

We consider the following assumptions throughout this section.

**Assumption 8:** The $q_{ij}^{(d)}$th communication channel from the C&C side filter to the plant side filter and the $q_{ip}^{(d)}$th communication channel from the plant side filter to the C&C side filter are secured, where $q_{ij}^{(d)} \in [1, \ldots, n_{ij}]$, i.e., rank($D_{ip}^{(d)}$) = $n_{ij} - 1$ and rank($D_{cpi}^{(d)}$) = $n_{ij} - 1$, for $i, j = 1, \ldots, N$. Moreover, all the communication channels among the nearby UIO-based detectors can be compromised, i.e., rank($D_{pp}^i$) = $n_i + p_i + p_i$.

**Assumption 9:** In the plant side filter (36), the C&C side filter (37), and the UIO-based detector (39) of $S_i$, only the random parameters $\delta_{pi}(t) L_i^e$, $\delta_{ci}(t) L_i^e$, and $\delta_{di}(t) L_i^e$ are unknown to the adversary, respectively, and the adversary knows the other parameters.

**Remark 11:** The random variables $\delta_{pi}(t)$, $\delta_{ci}(t)$, and $\delta_{di}(t)$ can have any arbitrary probability distributions.

**Proposition 7:** Consider Assumptions 8 and 9, the plant side filter (36), the C&C side filter (37), and the UIO-based detector (39). The residual signal $\text{res}_i^a(t) = y_{pi}(t) - C_i \hat{x}_i^e(t)$ in the $i$th subsystem is affected by actuator cyberattack signals $a_i^u(t)$ in $S_i$ and $a_i^d(t)$ in $S_j$ and malicious attack signals $a_{ip}^u(t)$ and $a_{ip}^{(d)}(t)$, for $j \in N_i$ and $r \in N_j$. Moreover, the generated $\text{res}_i^a(t)$ is decoupled from the sets of anomalous signals $\{a_i^u(t), f_i^j(t), f_i^j(t), a_i^d(t), f_i^j(t), f_i^j(t)\}$ and $\{a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t)\}$ in the sense of Definition 7 if for the error dynamics (38) and (40) and every $i = 1, \ldots, N$, the following conditions hold:

1) $T_i^e = I - H_i^e C_i$;
2) $(I - H_i^e C_i) F_i^e = 0$;
3) $(I - H_i^e C_i) F_i^e = 0$;
4) $L_i^e D_{ie}^{(d)} = 0$, $\tilde{L}_i^e D_{ie}^{(d)} = 0$, $L_i^e D_{ie}^{(d)} = 0$;
5) the triplet $(C_i, F_i^e, \tilde{t}_i^e)$ is input observable, where $\tilde{t}_i^e$ is the $q_{ij}^{(d)}$th column of $L_i^e$;
6) $F_i^e$ is Hurwitz;
7) $F_i^e$ is designed such that

$$
F_i^e(t) + F_i^{(j)}(t) = -Q_i^e(t)
$$

(42)

where $Q_i^e(t)$ is a symmetric positive definite matrix that satisfies $F_i^e I_r \leq Q_i^e(t)$, and $\beta_{cp}^j$ is a positive scalar;
8) and $K_i^{(1)} D_i^{(d)} = 0$.

**Proof:** Let $\ell = AA$. Considering the provided conditions in this proposition, the dynamics of errors $e_i^{AA}(t)$ and $e_j^{A}(t)$ can be derived in the following form, namely:

$$
\dot{e}_i^{AA}(t) = F_i e_i^{AA}(t) + (I - H_i^{AA} C_i) \sum_{j \in N_i} A_{ij} J_{ij}^{(i)} e_i^{AA}(t) - (I - H_i^{AA} C_i) \sum_{j \in N_i} A_{ij} J_{ij}^{(i)} e_i^{AA}(t)
$$

(43)

and

$$
\dot{e}_j^{A}(t) = F_j e_j^{AA}(t) + (I - H_j^{AA} C_j) \sum_{r \in N_j} A_{jr} J_{jr}^{(j)} e_j^{AA}(t) - (I - H_j^{AA} C_j) \sum_{r \in N_j} A_{jr} J_{jr}^{(j)} e_j^{AA}(t)
$$

(44)

where $r \in N_j$. From (43) and (44) it can be inferred that $\text{res}_i^{AA}(t) = C_i e_i^{AA}$ is affected by the actuator attack signals $a_i^u(t)$ and $a_i^d(t)$.

The remainder of the proof follows along similar lines to that of Propositions 1 and 5.

**Proposition 8:** Let Assumptions 8 and 9 hold and set $\ell = AA$. The residual signal $\text{res}_i^{AA}(t) = y_{pi}(t) - C_i \hat{x}_i^{AA}(t)$ generated using the plant side filter (36), the C&C side filter (37), and the UIO-based detector (39) in $S_i$ is affected by sensor cyberattack signals $a_{ip}^d(t)$ in the subsystem $i$ and $a_{ip}^d(t)$ in the $j$th subsystem and communication attack signals $a_{ip}^{(d)}(t)$ and $a_{ip}^{(d)}(t)$, for $j \in N_i$ and $r \in N_j$. Moreover, $\text{res}_i^{AA}(t)$ is decoupled from the sets of anomalous signals $\{a_i^u(t), f_i^j(t), f_i^j(t), a_i^d(t), f_i^j(t), f_i^j(t)\}$ and $\{a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t), a_i^{(d)}(t)\}$ in the sense of Definition 7 if for every $i = 1, \ldots, N$, $\{T_i^e, B_i^e\}$ is not affected by an anomalous signal $a_i^u(t)$ and $a_i^d(t)$.

**Proof:** The proof follows along similar lines to that of Propositions 5 and 7 and is omitted for sake of brevity.

**Proposition 9:** Let Assumption 8 holds and set $\ell = AA$. The residual signal $\text{res}_i^{AF}(t) = y_{pi}(t) - C_i \hat{x}_i^{AF}(t)$ that is
generated in the $i$th subsystem using the UIO-based detector (39) is affected by actuator faults $f_1^i(t)$ and $f_2^i(t)$ and malicious attack signals $a_{cp}^i(t)$ and $a_{pp}^i(t)$, for $j \in N_i$ and $r \in N_j$. Moreover, $\text{res}_{SA}^i$ is decoupled from the sets of anomalous signals $\{a_{u}^i(t), a_{c}^i(t), a_{f}^i(t), f_2^i(t)\}$ and $\{a_{cp}^i(t), a_{pc}^i(t), a_{pp}^i(t)\}$ in the sense of Definition 7, if the Conditions 1), 3), 4), and 6) of Proposition 7 hold and $L_i^j = 0$.

Proof: The proof follows along similar lines to that of Propositions 3 and 7.

**Proposition 10:** Consider Assumption 8 and let $\ell = SF$ for the modified UIO-based detector in (39). The generated residual signal $\text{res}_{SA}^i = y_{u}^i(t) - C_i^S x_i^S(t)$ in $S_i$ is affected by pseudo actuator faults $f_2^i(t)$ and $f_2^i(t)$ and communication channel attack signals $a_{cp}^i(t)$ and $a_{pp}^i(t)$, for $j \in N_i$ and $r \in N_j$. Furthermore, $\text{res}_{SA}^i$ is decoupled from the sets of anomalous signals $\{a_{u}^i(t), a_{c}^i(t), f_1^i(t), a_{r}^i(t), a_{i}^i(t), f_1^i(t)\}$ and $\{a_{cp}^i(t), a_{pc}^i(t), a_{pc}^i(t), a_{pp}^i(t)\}$ in the sense of Definition 7, if the Conditions 1), 2), 4), and 6) of Proposition 7 hold and $L_i^j = 0$.

Proof: The proof follows along similar lines to that of Propositions 3 and 7 and is omitted for sake of brevity.

**Theorem 2:** The residual signals $\text{res}_{SA}^i$ and $\text{res}_{SA}^j$ in Propositions 9 and 10, respectively, can be simultaneously generated to detect and isolate $f_1^i(t)$ and $f_2^i(t)$ if $F_1^i F_2 = 0$, for $i = 1, \ldots, N$.

Proof: The proof follows along similar lines to that of Theorem 1.

**Remark 12:** It should be pointed out that in Propositions 7-10, for detecting anomalies, i.e., faults and cyberattacks, in the neighboring subsystems, the matrix $H_i^\ell$ should be designed such that $(1 - H_i^\ell C_i)A_{ij} \neq 0$, for $i, j = 1, \ldots, N$ and $\ell \in \{SA, AA, SF, AF\}$.

**B. Nonsecure Communication Channels Among Two Side Filters and Nearby UIO-Based Detectors**

In this section, we consider the case where there is no secure communication channel among the plant side filter (36) and the C&C side filter (37), and vice versa. Hence, in the following, we investigate the performance of our proposed CFDAI methodology in the previous subsection under the following assumption.

**Assumption 10:** The adversary has access to all the communication channels among the two side filters and nearby UIO-based detectors, i.e., $\text{rank}(D_{cp}^j) = \text{rank}(D_{pp}^j) = n_1$ and $\text{rank}(D_{pp}^j) = n_j + p_{ij} + p_j$, for $i, j = 1, \ldots, N$.

Under Assumption 10, the Condition 4) in Proposition 7 cannot be satisfied. Hence, the impact of cyberattacks on the communication channels among the two side filters (37) and (36) and the nearby UIO-based detectors in (39) cannot be eliminated from the generated residuals in Propositions 7-10.

**Corollary 1 (Proposition 7):** Consider Assumptions 9 and 10. Since the Condition 4) in Proposition 7 cannot be satisfied, the generated residual $\text{res}^{SA} = y_{u}^i(t) - C_i^S x_i^S(t)$ cannot be decoupled from the sets of residual signals $U_a^i = \{a_{cp}^i(t), a_{pc}^i(t), a_{pp}^i(t)\}$ in $S_i$ and $U_d^j = \{a_{cp}^j(t), a_{pc}^j(t), a_{pp}^j(t)\}$ in the subsystem $S_j$ in the sense of Definition 7, for $j \in N_i$ and $r \in N_j$.

Proof: The proof follows along similar lines to that of Lemma 1 and Proposition 7.

**Corollary 2 (Proposition 8):** Let Assumptions 9 and 10 hold. The residual signal $\text{res}^{SA} = y_{u}^i(t) - C_i^S x_i^S(t)$ generated in Proposition 8 cannot be decoupled from the sets of attack signals $Y_a^i = \{a_{cp}^i(t), a_{pc}^i(t), a_{pp}^i(t)\}$ in the subsystem $i$ and $Y_a^j = \{a_{cp}^j(t), a_{pc}^j(t), a_{pp}^j(t)\}$ in the $j$th subsystem in the sense of Definition 7, where $j \in N_i$ and $r \in N_j$.

Proof: The proof follows along similar lines to that of Lemma 1 and Proposition 8 and is omitted for sake of brevity.

**Remark 13:** The Corollaries 1 and 2 indicate the importance of having one secure communication channel from the C&C side filter (37) to the plant side filter (36), and vice versa. Hence, under Assumption 10, although the residual signals $\text{res}^{SA} = y_{u}^i(t) - C_i^S x_i^S(t)$ can be used to detect and isolate actuator and sensor cyberattacks, respectively, they are also affected by the sets of anomalous signals $\{a_{cp}^i(t), a_{pc}^i(t), a_{pp}^i(t)\}$ and $\{a_{cp}^j(t), a_{pc}^j(t), a_{pp}^j(t)\}$.

**Remark 14:** In this article, two main centralized and distributed CAFDI methodologies have been developed and proposed. In Propositions 1-4, a centralized CAFDI methodology is proposed while it is assumed that there exists a set of secure communication channels to transmit information from the C&C side filter of the CPS (9) to its plant side UIO-based detector (12). Moreover, the designed CAFDI module in Propositions 1-4 only requires transmission of information from the C&C side to the plant side of the CPS. To eliminate our assumption on the number of secure communication channels, a modified version of our centralized CAFDI module is developed in Propositions 5 and 6 where one does not need to have any secure communication channel on the two sides. However, the proposed methodology in Propositions 5 and 6 requires one to transmit information both from the C&C side of the CPS to its plant side and from the plant side to its C&C side. Finally, in Propositions 7-10, our proposed CAFDI methodologies in previous sections have been extended to detect and isolate cyberattacks and faults in large-scale interconnected CPS through a distributed architecture. In Propositions 7-10, we assume that there exist one secure communication channel from the C&C side filter (37) to the plant side filter (36) and one secure communication channel from the plant side filter (36) to the C&C side filter (37). The proposed CAFDI methodology in Propositions 7-10 requires one to transmit information from the C&C side of the CPS to its plant side and from the plant side to its C&C side as well as the transmission of information among the nearby UIO-based detectors.

**V. CASE STUDY: FOUR AREA POWER NETWORK**

In this section, a HIL simulation is provided to demonstrate and verify the capabilities of our proposed methodologies. A four-area power network system is simulated by utilizing the OPAL-RT real-time simulator and 4 Raspberry Pis. As depicted in Fig. 5, the power system and all the plant side dynamics are simulated on the OPAL-RT simulator and the C&C side controllers and filters are deployed and simulated...
Fig. 5. Architecture of the HIL simulation for our proposed distributed CAFDI in the four area power network system. Black dashed and solid lines denote communication of data and physical couplings among subsystems, respectively.

Fig. 6. Implemented HIL simulation platform.

on the Raspberry Pis. Our HIL simulation setup is shown in Fig. 6. Similar to the distributed wide area monitoring systems (WAMSs), in Fig. 5, we have utilized phasor data concentrator (PDC) units which acquire, archive, exchange, and process data within each area [36], [37]. Furthermore, in [38] a methodology has been proposed which can be used to represent the IEEE New England 39-bus power system as a four area network.

The governing dynamics of the power system in the $i$th area is given by [39]

$$\dot{\delta}_i^v(t) = f^v_i(t)$$

$$T_{pi} \dot{f}^b_i(t) = -(f^b_i(t) - f^{nom}) + K_{pi}$$

$$\times \left( P_i(t) - P_{di} + \sum_{j \in N_i} V_i(t)V_j(t)B_{ij} \right.$$  

$$\left. \times \sin (\delta_i^v(t) - \delta_j^v(t)) \right)$$

$$T_{vi} \dot{V}_i(t) = \dot{E}_{li} - (X_{di} - X'_{di})B_{ii}V_i(t) - (X_{di} - X'_{di}) \sum_{j \in N_i} V_j(t)B_{ij} \cos (\delta_i^v(t) - \delta_j^v(t)))$$  

(45)

for $i = 1, \ldots, 4$. Moreover, the dynamics of the turbine and the governor can be expressed by [39]

$$T_t \dot{P}_t(t) = -P_t(t) + P_{gi}(t)$$

$$T_{gi} \dot{P}_{gi}(t) = -\frac{1}{R_i} \left( f^b_i(t) - f^{nom} \right) - P_{gi}(t) + u_i$$  

(46)

where the definition of parameters and their values in (45) and (46) are provided in Tables I and II, respectively. Also, $N_i$ is the set of neighborhoods of the subsystem $S_i$, for $i = 1, \ldots, 4$. In this case study, we have $N_1 = \{2, 4\}$, $N_2 = \{1, 3\}$, $N_3 = \{2, 4\}$, and $N_4 = \{1, 3\}$.

It should be noted that the nonlinear dynamics in (45) is used in the HIL simulation. However, to design our proposed CAFDI methodology for the interconnected large-scale CPS in Propositions 7–10, we have linearized the power system dynamics using Simulink “Model Linearizer” app. Consequently, by utilizing the linearized model, we design and implement a bank of two side filters and UIO-based detectors to detect and isolate cyberattacks and faults for the four area power network system in the HIL simulation platform.

A. HIL Simulation Results for the Four Area Power Network

In this case study, we develop a CAFDI methodology for a four area power network system under cyberattacks and faults. Each subsystem is connected to its neighboring subsystems through tie-lines with $B_{12} = -5.4$ p.u., $B_{23} = -5$ p.u.,

| Symbol | Description | Value |
|--------|-------------|-------|
| $\delta_i^v(t)$ | Voltage Angle | |
| $f^v_i(t)$ | Frequency | |
| $V_i(t)$ | Voltage | |
| $P_t(t)$ | Turbine Output Power | |
| $P_{gi}(t)$ | Governor Output Power | |
| $f^{nom}$ | Nominal Frequency | |
| $T_{pi}$ | Time Constant of the Generator | |
| $T_{vi}$ | Time Constant of the Turbine | |
| $T_{gi}$ | Time Constant of the Governor | |
| $K_{pi}$ | Governor Gain | |
| $R_i$ | Speed Regulation Coefficient | |
| $X_{di}$ | Direct Synchronous Reactance | |
| $X'_{di}$ | Direct Synchronous Transient Reactance | |
| $B_{ij}$ | Transmission Line Susceptance | |
| $u_i$ | Control Input to the Governor | |
| $\dot{E}_{li}$ | Constant Exciter Voltage | |
| $P_{di}$ | Unknown Power Demand | |

| Symbol | Value |
|--------|-------|
| $T_{pi}$ | 21 | 25 | 23 | 22 |
| $T_{vi}$ | 0.30 | 0.33 | 0.35 | 0.28 |
| $T_{gi}$ | 0.080 | 0.072 | 0.070 | 0.081 |
| $K_{pi}$ | 5.54 | 7.41 | 6.11 | 6.22 |
| $R_i$ | 120 | 112.5 | 115 | 118.5 |
| $X_{di}$ | 2.5 | 2.7 | 2.6 | 2.8 |
| $X'_{di}$ | 1.85 | 1.84 | 1.86 | 1.83 |
| $B_{ij}$ | 0.25 | 0.24 | 0.26 | 0.23 |
| $u_i$ | -13.6 | -12.9 | -12.3 | -12.3 |
| $\dot{E}_{li}$ | 1 | 1 | 1 | 1 |
| $P_{di}$ | 0.01 | 0.015 | 0.012 | 0.014 |
\( B_{34} = -4.5 \) p.u., and \( B_{14} = -5.2 \) p.u., with the base power of 1000 MW.

In this case study, we consider that the actuator of each subsystem as well as the first and the third sensors of all the subsystems are under faulty conditions. Furthermore, all the input and output channels of all the subsystems are compromised by adversaries. As depicted in Fig. 5, each subsystem is equipped with a bank of plant side filters given by (36), C&C side filters in (37), and UIO-based detectors provided in (39) that are designed according to Propositions 7–10. As per Assumption 8, we assume that there exists one secured communication channel from (37) to (36) and one secured communication channel from (36) to (37).

**Scenario 1 (Covert Attacks):** In this scenario, a covert attack on the subsystem \( S_1 \) is considered. The covert attack starts at \( t = 50 \) (s) and ends at \( t = 300 \) (s). As shown in Fig. 7 the impact of both the actuator and sensor attacks on \( S_1 \) can be seen in the generated residual of \( S_1 \) and nearby subsystems which are \( S_2 \) and \( S_4 \). Hence, the covert cyberattack on the subsystem \( S_1 \) is detected by this subsystem and its neighboring subsystems.

**Scenario 2 (Faults):** Actuator and sensor faults are injected to the subsystem \( S_1 \). The subsystem \( S_1 \) is under an actuator fault from \( t = 60 \) (s) to \( t = 300 \) (s), and a sensor fault starting from \( t = 150 \) (s) to \( t = 350 \) (s). As depicted in Fig. 8, \( S_1 \) has been able to detect its local actuator fault, but the neighboring subsystems, i.e., \( S_2 \) and \( S_4 \), have not detected the actuator fault. This is due to the linearization that was made for (45) which has resulted in having \((I - H_{A}^{AF} C_1)A_{1j} = 0\), for \( j = 2 \) and 4 (see Remark 12). However, the sensor fault is successfully detected and isolated in the subsystems \( S_1 \), \( S_3 \), and \( S_4 \).

**Scenario 3 (Simultaneous Injection of Cyberattacks and Faults):** In this scenario, actuator and sensor cyberattacks as well as actuator and sensor faults are simultaneously injected to the subsystem \( S_1 \). The subsystem \( S_1 \) is under an actuator attack starting from \( t = 50 \) (s) to \( t = 220 \) (s), a sensor attack starting from \( t = 100 \) (s) to \( t = 250 \) (s), an actuator fault from \( t = 150 \) (s) to \( t = 300 \) (s), and a sensor fault starting from \( t = 200 \) (s) to \( t = 350 \) (s). As depicted in Fig. 9, all the anomalies are successfully detected and isolated.

---

**Fig. 7.** Detection of covert cyberattack on the subsystem \( S_1 \).

**Fig. 8.** Detection and isolation of faults in the subsystem \( S_1 \) and nearby subsystems.

**Fig. 9.** Detection and isolation of simultaneous cyberattacks and faults in the subsystem \( S_1 \) and nearby subsystems.

---

**B. Performance Evaluation**

A confusion matrix analysis [40] is employed to evaluate the performance of our proposed CAFDI methodology for the four area power network. Given a classifier and its corresponding instances, four possible outcomes are obtained as: 1) True Positive (TP), if the instance is positive and is truly classified as positive; 2) False Negative (FN), if the instance is positive and incorrectly classified as negative; 3) True Negative (TN), if the instance is negative and correctly classified as negative; and 4) False Positive (FP), if the instance is negative and incorrectly classified as positive [40]. Based on the possible outcomes, true positive rate (TPR) and false positive rate (FPR) can be used as performance metrics and measures, where TPR = TP/(TP + FN) and FPR = FP/(FP + TN). In particular, the receiver operating characteristic (ROC) curve which shows the TPR versus FPR for various threshold levels is also utilized in this section.

The ROC curves for \( S_1 \), \( S_2 \), and \( S_4 \) are depicted in Fig. 10. Moreover, as illustrated in the previous section, cyberattacks and faults on \( S_1 \) cannot be detected in \( S_3 \), therefore, the ROC curve for the third subsystem is not provided. It can be seen in Fig. 10 that for the case of actuators cyberattacks, sensors cyberattacks, and sensor faults \( S_1 \), \( S_2 \), and \( S_4 \) have high TPR. Moreover, the neighboring subsystems \( S_2 \) and
In this article, the problem of simultaneous detection and isolation of machine-induced faults and intelligent malicious adversarial cyberattacks has been studied. Centralized and distributed methodologies based on the CPSs two side filters and UIO-based detectors have been proposed and developed. In both methodologies, a bank of filters along with UIO-based detectors are designed on the plant side and a bank of filters was implemented on the C&C side of the CPS. In case of the proposed distributed CAFDI methodology, the UIO-based detector of each subsystem communicates information with UIO-based detectors in the nearby subsystems. Hence, under certain conditions, each subsystem can detect and isolate its cyberattacks and faults as well as anomalies in its nearby subsystems. Using the proposed centralized and distributed strategies, one is capable of simultaneously detecting machine-induced actuator and sensor faults as well as undetectable cyberattacks, such as covert and zero dynamics attacks, and detectable cyberattacks, such as false data injection attacks. In our future work, we will consider CPS that are represented by nonlinearities in their dynamics.

VI. CONCLUSION

REFERENCES

[1] 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.
[2] 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.
[3] S. K. Khaitan and J. D. McCalley, “Design techniques and applications of cyberphysical systems: A survey,” IEEE Syst. J., vol. 9, no. 2, pp. 350–365, Jun. 2015.
[4] Y. Li, P. Zhang, L. Zhang, and B. Wang, “Active synchronous detection of deception attacks in microgrid control systems,” IEEE Trans. Smart Grid, vol. 8, no. 1, pp. 373–375, Jan. 2017.
[5] F. Miao, M. Pajic, and G. J. Pappas, “Stochastic game approach for replay attack detection,” in Proc. 52nd IEEE Conf. Decis. Control, Dec. 2013, pp. 1854–1859.
[6] M. Taheri, K. Khorasani, I. Shames, and N. Meskin, “Undetectable cyber attacks on communication links in multi-agent cyber-physical systems,” in Proc. 59th IEEE Conf. Decis. Control (CDC), Dec. 2020, pp. 3764–3771.
[7] Z. Zhao, Y. Yang, Y. Li, and R. Liu, “Security analysis for cyber-physical systems under undetectable attacks: A geometric approach,” Int. J. Robust Nonlinear Control, vol. 30, no. 11, pp. 4359–4370, Jul. 2020.
[8] A. Baniamierian and K. Khorasani, “Security index of linear cyber-physical systems: A geometric perspective,” in Proc. 6th Int. Conf. Control, Decis. Inf. Technol. (CoDIT), Apr. 2019, pp. 391–396.
[9] A. Baniamierian, K. Khorasani, and N. Meskin, “Determination of security index for linear cyber-physical systems subject to malicious cyber attacks,” in Proc. IEEE 58th Conf. Decis. Control (CDC), Dec. 2019, pp. 4507–4513.
[10] R. M. G. Ferrari and A. M. H. Teixeira, “A switching multiplicative watermarking scheme for detection of stealthy cyber-attacks,” IEEE Trans. Autom. Control, vol. 66, no. 6, pp. 3558–3573, Jun. 2021.
[11] M.-A. Massoumnia, “A geometric approach to the synthesis of failure detection filters,” IEEE Trans. Autom. Control, vol. AC-31, no. 9, pp. 839–846, Sep. 1986.
[12] S. H. Zad and M.-A. Massoumnia, “Generic solvability of the failure detection and identification problem,” Automatica, vol. 35, no. 5, pp. 887–893, May 1999.
[13] J. Chen, R. J. Patton, and H.-Y. Zhang, “Design of unknown input observers and robust fault detection filters,” Int. J. Control, vol. 63, no. 1, pp. 85–105, Jan. 1996.
[14] J. Wünnenberg and P. Frank, “Sensor fault detection via robust observers,” in System Fault Diagnostics, Reliability and Related Knowledge-Based Approaches. Dordrecht, The Netherlands: Reidel, 1987, pp. 147–160.
[15] K. K. N. Tudoroiu, “Fault detection and diagnosis for satellite’s attitude control system (ACS) using an interacting multiple model (IMM) approach,” in Proc. IEEE Conf. Control Appl. (CCA), Aug. 2005, pp. 1287–1292.
[16] N. Meskin, E. Naderi, and K. Khorasani, “A multiple model-based approach for fault diagnosis of jet engines,” IEEE Trans. Control Syst. Technol., vol. 21, no. 1, pp. 254–262, Jan. 2013.
[17] M. Davoodi, N. Meskin, and K. Khorasani, “Simultaneous fault detection and consensus control design for a network of multi-agent systems,” Automatica, vol. 66, pp. 185–194, Apr. 2016.
[18] I. Shames, A. M. H. Teixeira, H. Sandberg, and K. H. Johansson, “Distributed fault detection for interconnected second-order systems,” Automatica, vol. 47, no. 12, pp. 2757–2764, Dec. 2011.
[19] J. Gertler, “Fault detection and isolation using parity relations,” Control Eng. Pract., vol. 5, no. 5, pp. 653–661, May 1997.
[20] R. J. Patton and J. Chen, “A review of parity space approaches to fault diagnosis,” IFAC Proc. Volumes, vol. 24, no. 6, pp. 65–81, Sep. 1991.
[21] Y. Mo, S. Weerakkody, and P. Zhang, “Detection of covert attacks and zero dynamics attacks in cyber-physical systems,” in Proc. Amer. Control Conf. (ACC), Jul. 2016, pp. 302–307.
[22] A. Barbouni, H. Rezaee, F. Boem, and T. Parisini, “Detection of covert cyber-attacks in interconnected systems: A distributed model-based approach,” IEEE Trans. Autom. Control, vol. 65, no. 9, pp. 3728–3741, Sep. 2020.
[23] Y. Mo, S. Weerakkody, and B. Sinopoli, “Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs,” IEEE Control Syst. Mag., vol. 35, no. 1, pp. 93–109, Feb. 2015.
[24] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, “Revealing stealthy attacks in control systems,” in Proc. 50th Ann. Allerton Conf. Commun., Control, Comput. (Allerton), Oct. 2012, pp. 1806–1813.
[25] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, “Coding schemes for securing cyber-physical systems against stealthy data injection attacks,” IEEE Trans. Control Netw. Syst., vol. 4, no. 1, pp. 106–117, Mar. 2017.
[26] S. Weerakkody and B. Sinopoli, “Detecting integrity attacks on control systems using a moving target approach,” in Proc. 54th IEEE Conf. Decis. Control (CDC), Dec. 2015, pp. 5821–5826.
[27] C. Schellenberger and P. Zhang, “Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system,” in Proc. IEEE 56th Ann. Conf. Decis. Control (CDC), Dec. 2017, pp. 1374–1379.
[28] K. Zhang, C. Kelliris, M. M. Polycarpou, and T. Parisini, “Discrimination between replay attacks and sensor faults for cyber-physical systems via event-triggered communication,” Eur. J. Control, vol. 62, pp. 47–56, Nov. 2021.
S. Chockalingam, W. Pieters, A. Teixeira, and P. van Gelder, “Bayesian network model to distinguish between intentional attacks and accidental technical failures: A case study of floodgates,” Cybersecurity, vol. 4, no. 1, pp. 1–19, Dec. 2021.

H. L. Trentelman, A. A. Stoorvogel, and M. Hautus, Control Theory for Linear Systems. Cham, Switzerland: Springer, 2012.

M. A. Massoumnia, C. G. Verghe, and A. S. Willsky, “Failure detection and identification,” IEEE Trans. Autom. Control, vol. 34, no. 3, pp. 316–321, Mar. 1989.

M. Mola, N. Meskin, K. Khorasani, and A. Massoud, “Distributed event-triggered consensus-based control of DC microgrids in presence of DoS cyber attacks,” IEEE Access, vol. 9, pp. 54009–54021, 2021.

A. J. Gallo, M. S. Turan, F. Boem, T. Parisini, and G. Ferrari-Trecate, “A distributed cyber-attack detection scheme with application to DC microgrids,” IEEE Trans. Autom. Control, vol. 65, no. 9, pp. 3800–3815, Sep. 2020.

W. M. Wonham, “Linear multivariable control,” in Optimal Control Theory and Its Applications. Berlin, Germany: Springer, 1974, pp. 392–424.

H. K. Khalil and J. W. Grizzle, Nonlinear Systems. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

R. Gore and M. Kande, “Analysis of wide area monitoring system architectures,” in Proc. IEEE Int. Conf. Ind. Technol. (ICIT), Mar. 2015, pp. 1269–1274.

S. Nabavi, J. Zhang, and A. Chakrabortty, “Distributed optimization algorithms for wide-area oscillation monitoring in power systems using interregional PMU-PDC architectures,” IEEE Trans. Smart Grid, vol. 6, no. 5, pp. 2529–2538, Sep. 2015.

S. Nabavi and A. Chakrabortty, “Topology identification for dynamic equivalent models of large power system networks,” in Proc. Amer. Control Conf., Jun. 2013, pp. 1138–1143.

S. Trip, M. Cucuzzella, C. De Persis, A. van der Schaft, and A. Ferrara, “Passivity-based design of sliding modes for optimal load frequency control,” IEEE Trans. Control Syst. Technol., vol. 27, no. 5, pp. 1893–1906, Sep. 2019.

T. Fawcett, “An introduction to ROC analysis,” Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, Jun. 2006.

Iman Shames (Member, IEEE) received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 2006, and the Ph.D. degree in engineering from The Australian National University, Canberra, ACT, Australia, in 2011. He is a Professor of mechatronics engineering with the School of Engineering, The Australian National University. Before joining ANU, he was a McKenzie Fellow and an Associate Professor with the Department of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, VIC, Australia, from 2012 to 2014 and from 2020 to 2021, respectively; a Senior Lecturer, from 2014 to 2019. Previously, he was an ACCESS Post-Doctoral Researcher with the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden. His current research interests include, but are not limited to, decision making under uncertainty, optimization theory and its application in control and estimation, mathematical systems theory of cyber-physical systems, and security and privacy.

Dr. Shames is serving as an Associate Editor for IEEE Transactions on Control of Network Systems.

Nader Meskin (Senior Member, IEEE) received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1998, the M.Sc. degree from the University of Tehran, Tehran, in 2001, and the Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2008. He was a Post-Doctoral Fellow with Texas A&M University at Qatar, Doha, Qatar, from January 2010 to December 2010. He is currently a Professor with Qatar University, Doha, and an Adjunct Associate Professor with Concordia University. He has published more than 270 refereed journal and conference papers. His research interests include FDI, multiagent systems, active control for clinical pharmacology, and cyber-security of industrial control systems.

Dr. Meskin is serving as a member for Editorial Boards for IEEE Transactions on Cybernetics.