Hierarchical Cyber-Attack Detection in Large-Scale Interconnected Systems

Keijzer, T.; Gallo, A.J.; Ferrari, Riccardo M.G.

DOI
10.1109/CDC51059.2022.9993072

Publication date
2022

Document Version
Final published version

Published in
Proceedings of the IEEE 61st Conference on Decision and Control (CDC 2022)

2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Hierarchical Cyber-Attack Detection in Large-Scale Interconnected Systems

Twan Keijzer*, Alexander J. Gallo* and Riccardo M.G. Ferrari*

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9993072

Abstract— **In this paper we present a hierarchical scheme to detect cyber-attacks in a hierarchical control architecture for large-scale interconnected systems (LSS). We consider the LSS as a network of physically coupled subsystems, equipped with a two-layer controller: on the local level, decentralized controllers guarantee overall stability and reference tracking; on the supervisory level, a centralized coordinator sets references for the local regulators. We present a scheme to detect attacks that occur at the local level, with malicious agents capable of affecting the local control. The detection scheme is computed at the supervisory level, requiring only limited exchange of data and model knowledge. We offer detailed theoretical analysis of the proposed scheme, highlighting its detection properties in terms of robustness, detectability and stealthiness conditions.**

I. Introduction

Modern engineering systems, ranging from large-scale infrastructure like electrical grids, water distribution and traffic networks, as well as industrial plants and consumer goods, have an increasing penetration of distributed computational resources, and a heavy reliance on communication networks. This improves performance and efficiency of these systems, and has led to the definition of *cyber-physical systems* (CPS) [1] as an analytical framework. Although the integration of these "cyber" resources has great benefits, it also leads to the exposure to malicious tampering, as has been made evident in recent years by some high profile cases of cyber-attacks [2], [3]. Because many of these systems are safety-critical [4], methods have been developed over the past decade to detect, isolate and mitigate attacks in CPS [5], [6].

CPS are often also large-scale systems (LSS) [7], [8], i.e., they require a large number of states to be described, and are spatially distributed over large areas. As such, centralized control architectures, with a single regulator managing the inputs for the entire system, are not feasible. Thus, non-centralized control architectures have been developed, which rely on partitioning the LSS into *subsystems* [7], each of which is physically interconnected with neighbouring subsystems. These control architectures can be further classified as decentralized, distributed, and hierarchical, all of which have attracted extensive literature (see for example the surveys [9], [10], and references therein).

Non-centralized monitoring architectures have been proposed for both fault diagnosis [11], [12], [13], [14], [15], [16], [17] and cyber-attack detection [18], [19], [20], [21], [22], predominantly on distributed and decentralized architectures. Furthermore, there are however application based papers, such as [23] addressing energy theft, where implicitly a hierarchical framework is used for cyber-attack detection.

On the other hand, here we focus on hierarchical architectures. In hierarchical control, the computational advantages of distributed or decentralized controllers are blended with the coordination capability of a centralized, *supervisory*, layer [9]. Hierarchical control indeed appears naturally in large-scale interconnected systems, as it is an architecture that allows for multiple degrees of complexity and coordination to be integrated.

In this paper, we propose a hierarchical cyber-attack detection scheme which leverages the physical coupling between local subsystems to detect attacks. By computing two estimates, at the local and supervisory level, sufficient redundancy is introduced to perform diagnoses. For this detection scheme:

a. cyber-attacks fully compromising one or more local controllers can be detected at the supervisory level;
b. the supervisory level requires only a reduced order representation of the subsystem dynamics for detection, allowing for reduced computational overhead;
c. a thorough theoretical analysis of the detection properties is presented, including robustness, guaranteed attack detectability conditions, and existence conditions for locally and globally stealthy attacks.

The use of physical coupling to generate the necessary redundancy for detection has been proven beneficial in distributed cyber-attack diagnosis schemes [21], [22], and indeed we show that it is a critical aspect of our proposed hierarchical diagnoser. Here, subsystem model knowledge and measurements are used to define a local estimate of the physical coupling, which is then compared to a supervisory estimate computed from global knowledge.

We make use of a set-based detection scheme, which has a rich history as a detection method in the FDI literature, as can be seen in, e.g., [16], [24], [25], [26], [27], [28], where this list does not have the pretence of being exhaustive. In this paper, but without loss of generality, we adopt constrained zonotopes [26], which are proven to offer numerical advantages with respect to other set representations.

The rest of the paper is structured as follows: in Section II we introduce the problem formulation, describing the dynamics of the large-scale interconnected systems, and formalizing the attacks considered in this paper. Then, in Section III, we outline a hierarchical controller. In Section IV, we present our proposed hierarchical scheme for estimating the physical

*Delft Centre for Systems and Control, Delft University of Technology, The Netherlands. {t.keijzer,a.j.gallo,r.ferrari}@tudelft.nl

This paper has been partially supported by the AIMWIND project, which is funded by the Research Council of Norway under grant no. 312486.

interconnection between subsystems, and give the definition of our detection test. Following this, in Section V, we offer theoretical analysis for our proposed method. Finally, in Section VI we provide some numerical results, and in Section VII we offer concluding remarks.

Notation: For a matrix A, $\|A\|$ represents the induced Euclidian norm of A, and $\ker A$ its right null-space. $\mathrm{col}_{i\in\mathcal{I}}[x_i]$ and $\mathrm{diag}_{i\in\mathcal{I}}[x_i]$ denote respectively the column and block-diagonal concatenation of vectors or matrices $x_i\ \forall\ i\in\mathcal{I}\triangleq\{1,\dots,N\}$. Given matrices $A_{ij}, i,j\in\mathcal{I}$ of appropriate dimensions, $A=[A_{ij}]$ denotes the block matrix with A_{ij} in the i,j-th block. For sets $\mathcal{A}\subseteq\mathbb{R}^n$ and $\mathcal{B}\subseteq\mathbb{R}^n$ we denote with $\oplus$ the Minkowski sum, $\mathcal{A}\oplus\mathcal{B}\triangleq\{x\in\mathbb{R}^n : x=a+b, \forall a\in\mathcal{A}, b\in\mathcal{B}\}$; with $\mathcal{A}\ominus\mathcal{B}$ the erosion, or Pontryagin difference, of $\mathcal{A}$ by $\mathcal{B}$, $A\ominus B\triangleq\{\zeta\in\mathbb{R}^n|\zeta+b\in A, \forall b\in\mathcal{B}\}$, and therefore $\mathcal{A}\ominus\mathcal{A}=0$ [29]. The cartesian product of two sets $\mathcal{A}$ and $\mathcal{B}$ is defined as $\mathcal{A}\times\mathcal{B}$. Additionally, for a set $\mathcal{A}$, $\mathrm{Vol}(\mathcal{A})$ denotes its volume and $\mathcal{A}'$ its set complement. Furthermore, constrained zonotope $\mathcal{Z}\subseteq\mathbb{R}^n$ is defined as $\mathcal{Z}\triangleq\{\zeta\in\mathbb{R}^n : \zeta=c+G\beta,\ \|\beta\|_\infty\leq 1, A\beta=b\}$. Efficient definitions of set operations with constrained zonotopes can be found in [26]. Finally, with $\bigotimes_{i\in\mathcal{I}}A_i$ we intend $A_1\times A_2\times\cdots\times A_N$.

II. Problem Statement

We consider a linear time-invariant *Large-Scale System* (LSS) which is partitioned into N physically coupled subsystems $\mathcal{S}_i, i\in\{1,\dots,N\}\triangleq\mathcal{N}$. The dynamics of each subsystem is written as

$$\begin{cases} x_i(k+1)=A_{ii}x_i(k)+B_iu_i(k)+d_i(k)\,, \\ d_i(k)=\sum_{j\in\mathcal{N}_i}A_{ij}x_j(k) \end{cases} \tag{1}$$

where $x_i\in\mathbb{R}^{n_i}, u_i\in\mathbb{R}^{m_i}$ are the subsystem state and control input. The term $d_i(k)$ accounts for the physical coupling between subsystems, where $\mathcal{N}_i\triangleq\{j\in\{1,2,\dots,N\} : \partial x_i/\partial x_j\neq 0\}\subseteq\mathcal{N}$ is the set of *neighbors* of $\mathcal{S}_i$, i.e., those subsystems which physically influence the dynamics of $\mathcal{S}_i$.

Assumption 1: For all $i\in\mathcal{N}$, (A_{ii},B_i) is controllable. ◁

We suppose the LSS is regulated via a hierarchical control architecture, composed of two layers, as shown in Figure 1, with the following characteristics, detailed in Section III:

- locally, each subsystem is regulated by a decentralized controller $\mathcal{C}_i^\ell$. This guarantees LSS stability and is capable of tracking a suitably defined reference r_i;
- at the supervisory level, a controller $\mathcal{C}^s$ is designed to provide appropriate references r_i to the local controllers, thus providing coordination for the LSS.

A. Cyber-attack vulnerability

In this work, we consider a cyber-attack carried out by an agent capable of fully compromising a subset of the subsystems and their controllers. In order to clearly define the problem we address, we introduce the following assumption.

Assumption 2: An attacker may attack subsystems with indexes $\mathcal{I}_a\subseteq\mathcal{N}$ from some time $k_a\geq 0$. ◁

The considered attack is modelled as $u_i=u_i^0+a_i^u$ where u_i^0 is the healthy input as obtained by $\mathcal{C}_i^\ell$ and a_i^u is, without loss of generality, an additive attack.

The control architecture considered in this paper requires a communication network that links the supervisory controller $\mathcal{C}^s$ to all the local controllers $\mathcal{C}_i^\ell$. Thus, $\mathcal{C}^s$ represents a single point of failure, and if it were compromised by a malicious agent, it could steer the entire LSS to any desired operating condition. Given this premise, we suppose that the hardware and software of the supervisory controller are suitably designed to give a higher degree of protection, and therefore $\mathcal{C}^s$ cannot be subject to attacks.

B. Hierarchical Cyber-Attack Detection

Let us, before introducing the hierarchical control architecture considered in this paper, briefly give an overview of our proposed method. The detection architecture relies on comparing two estimates of d_i: one of the estimates is computed locally for each subsystem in $\mathcal{N}$, based on local model knowledge and measurements; a second estimate is computed at the supervisory level, using information on the references followed by each subsystem. During operation, the local estimate is transmitted to the supervisory level, where a detection test is performed. Detection is then performed by comparing two sets bounding the nominal coupling d_i given the local and supervisory estimates and their respective estimation errors. The proposed method allows these sets to be constructed at the supervisory level in a computationally efficiently way using only limited model knowledge of the LSS.

III. Hierarchical Control

Let us now describe the hierarchical controllers regulating the LSS. We stress again that this control architecture requires a communication network to exchange information between each local controller $\mathcal{C}_i^\ell$ and the supervisory controller $\mathcal{C}^s$, whilst not communicating amongst each other. The control architecture is represented in Figure 1.

A. Local controllers $\mathcal{C}_i^\ell$

We consider a decentralized $\mathcal{C}_i^\ell$ designed to ensure stability of the LSS, while locally tracking a reference. The reference $r_i\in\mathbb{R}^{q_i}$ is set by the supervisory controller $\mathcal{C}^s$, and follows the dynamics:

$$r_i(k+1)=S_ir_i(k), \tag{2}$$

with S_i such that its eigenvalues have modulus no smaller than one [30]. Thus, defining the output tracking error $e_i\in\mathbb{R}^{p_i}$, $e_i=C_ix_i+Q_ir_i$, the controller's tracking objective is

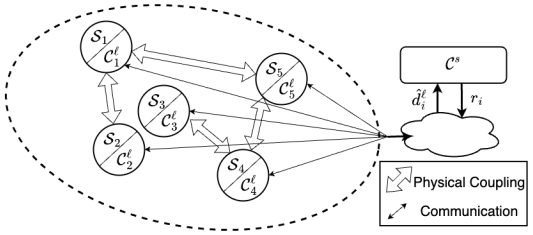

Fig. 1. The considered hierarchical control and detection architecture.

to define u_i such that nominally, i.e., when the system is not under attack, $\lim_{k\to\infty} e_i(k) = 0$. Specifically, we focus on $\mathcal{C}_i^\ell$ capable of solving the full-information regulator problem [30].

Assumption 3: For all $i \in \mathcal{N}$, the full state x_i is measured by the controller $\mathcal{C}_i^\ell$ for regulation purposes. ◁

Remark 1: Assumption 3, although potentially limiting, is introduced here to simplify the analysis of the controllers C_i^ℓ, which is not the primary focus of this paper. ◁

Assumption 4: For each subsystem $\mathcal{S}_i, i \in \mathcal{N}$, $\ker C_i \subseteq \ker A_{ji}$, for all $\mathcal{S}_j$ such that $i \in \mathcal{N}_j$. ◁

We consider the control law

$$\mathcal{C}_i^\ell : \quad u_i^0(k) = K_i x_i(k) + L_i r_i(k)\,, \tag{3}$$

where K_i and L_i satisfy the output regulation problem [30]. Specifically, K_i is designed such that, while $r_i = 0, \forall i \in \mathcal{N}$, the closed-loop LSS dynamics is asymptotically stable (for methods to design K_i, see [7], [8] and references therein). Before defining L_i, we introduce the following.

Assumption 5: For each subsystem $i \in \mathcal{N}$, the matrices A_i, B_i, C_i, Q_i, S_i are such that there exist $\Pi_i \in \mathbb{R}^{n_i \times q_i}$ and $\Gamma_i \in \mathbb{R}^{m_i \times q_i}$ such that:

$$\Pi_i S_i = A_i \Pi_i + B_i \Gamma_i \tag{4a}$$

$$0 = C_i \Pi_i + Q_i \tag{4b}$$

holds. ◁

Assumption 5 guarantees that the so-called regulator equations (4) can be solved, and thus L_i can be defined as

$$L_i = \Gamma_i - K_i \Pi_i\,, \tag{5}$$

which guarantees that $x_i \to \Pi_i r_i$ as $k \to \infty$ [30], supposing $d_i = 0, \forall i \in \mathcal{N}$. Note that satisfaction of (4b) guarantees that $\epsilon_i \triangleq x_i - \Pi_i r_i = 0$ implies $C_i \epsilon_i = e_i = 0$. On the other hand, for $d_i \neq 0$, we have that the dynamics of the state tracking error $\epsilon_i(k) = x_i(k) - \Pi_i r_i(k)$ are

$$\begin{aligned} \epsilon_i(k+1) &= A_{ii} x_i(k) + B_i u_i^0(k) + d_i(k) - \Pi_i S_i r_i(k) \\ &\overset{(3),(4a)}{=} A_{ii}^{cl}(x_i(k) - \Pi_i r_i(k)) + d_i(k)\,, \end{aligned} \tag{6}$$

where $A_{ii}^{cl} \triangleq A_{ii} + B_i K_i$. Given the stability of A_{ii}^{cl}, by design of K_i, ϵ_i is bounded for bounded d_i.

B. Supervisory-level controller $\mathcal{C}^h$

Having presented the local decentralized tracking controller $\mathcal{C}_i^\ell$, we can briefly discuss the design of $\mathcal{C}^s$. As previously stated, the objective of $\mathcal{C}^s$ is to design $r_i, \forall i \in \mathcal{N}$ such that some level of coordination between subsystems is possible. Although the specific design of r_i is dependent on the type of application considered, and is out of the scope of this paper, we introduce some basic characteristics that must be included in its design. We suppose that $\mathcal{C}^s$ defines r_i as a piecewise constant signal. This in turn implies that the reference dynamics in (2) are $r_i(k+1) = r_i(k)$ for almost all k, and therefore that $S_i = I_{p_i}, \forall i \in \mathcal{N}$; furthermore, set $Q_i = -I_{p_i}, \forall i \in \mathcal{N}$. This definition of S_i still allows for the references to be changed at discrete time instances. However, depending on the rate of convergence of A_{ii}^{cl}, it is important to specify a minimum time between switching times, and a maximum step in r_i [31].

IV. Hierarchical Attack Detection

The hierarchical cyber-attack detection scheme presented in this paper uses the physical interconnection between subsystems to perform detection at the supervisory level. To this end two estimates of this physical interconnection are computed. The so-called *local estimate* depends on local measurements as well as local model information, and is calculated at each local subsystem and communicated to the supervisory level. The so-called *supervisory estimate* is calculated at the supervisory level and depends on knowledge of the interconnection and a simplified model of the local dynamics. These two estimates, along with their estimation uncertainties, are compared at the supervisory level to detect cyber-attacks using a set-based approach.

A. Supervisory Estimate

The supervisory estimate of the physical interaction between all subsystems is computed as

$$\hat{d}^s = A_c \Pi r\,, \tag{7}$$

where $A_c = A - \mathrm{diag}_{i\in\mathcal{N}}[A_{ii}]$, $A = [A_{ij}]$, $\Pi = \mathrm{diag}_{i\in\mathcal{N}}[\Pi_i]$ and $r = \mathrm{col}_{i\in\mathcal{N}}[r_i]$. Thus, the estimation error is $e^s \triangleq d - \hat{d}^s = A_c \epsilon$, where $d = \mathrm{col}_{i\in\mathcal{N}}[d_i]$ and $\epsilon = \mathrm{col}_{i\in\mathcal{N}}[\epsilon_i]$. Thus, in nominal conditions the error can be bounded as $e^s \in \mathcal{E}^s$, with:

$$\mathcal{E}^s(k) = A_c \mathcal{E}(k), \tag{8}$$

where $\mathcal{E} = \bigotimes_{i\in\mathcal{N}} \mathcal{E}_i$, and

$$\mathcal{E}_i(k) \triangleq \{\zeta \in \mathbb{R}^n | \|\zeta\| \leq \bar{\epsilon}_i(k)\}\,, \tag{9}$$

where $\bar{\epsilon}_i(k)$ guarantees $\|\epsilon_i(k)\| \leq \bar{\epsilon}_i(k), \forall k < k_a$, such that $\epsilon_i(k) \in \mathcal{E}_i(k), \forall k < k_a$: the trajectory of $\bar{\epsilon}_i$ is the result of the dynamics

$$\begin{aligned} \bar{\epsilon}_i(k+1) = b_i \bar{\epsilon}_i(k) &+ \|\hat{d}_i^s(k)\| + \|\mathcal{E}_i^s(k)\| \\ &+ \|\Pi_i\| \|r_i(k+1) - r_i(k)\|\,, \end{aligned} \tag{10}$$

which are defined bounding (6) via the triangle inequality, where $b_i \in [0,1]$ is defined such that $\|A_{ii}^{cl\,k} \epsilon_i(0)\| \leq b_i^k \bar{\epsilon}_i^0$ for all $i \in \mathcal{N}$, $\bar{\epsilon}_i^0$ is an appropriately defined initial bound, $\hat{d}_i^s(k)$ are the components of $\hat{d}^s(k)$ relating to $\mathcal{S}_i$, and $\mathcal{E}_i^s$ is the projection of $\mathcal{E}$ onto the space relating to $\mathcal{S}_i$. Furthermore, $\|r_i(k+1) - r_i(k)\|$ is added to bound the effect of reference changes. By using a bound on the norm of ϵ_i, only the rate of convergence b_i, is needed at the supervisory level. Then, via (7), (8) the *supervisory estimation set* is defined

$$\mathcal{D}^s \triangleq \hat{d}^s \oplus \mathcal{E}^s\,. \tag{11}$$

Thus, $d(k) \in \mathcal{D}^s$ holds by construction for all $k < k_a$. Note that $\mathcal{D}^s$ only depends on r and is therefore not affected by the considered cyber-attacks. Furthermore $\mathcal{E}_i$ is represented as a hyper-sphere in n_i-dimensions. As such sets are not closed under matrix multiplication as done in (8) we define a constrained zonotope $\bar{\mathcal{E}}_i$ that encloses $\mathcal{E}_i$

and use this for further calculation. An exact representation of $\mathcal{E}_i$ as a constrained zonotope requires a number of generators approaching infinity which is computationally infeasible, while a hyper-cube with n_i generators may not be sufficiently accurate. Therefore, the representation of $\bar{\mathcal{E}}_i$ should be constructed to balance computational requirements with accuracy.

Remark 2: Computing enclosing sets can be computationally intensive, however $\mathcal{E}_i$ is always a hyper-sphere centered in 0. Thus, an enclosing set can be computed once off-line for the unit hyper-sphere, to only be re-scaled on-line. ◁

B. Local Estimate

The local estimator in each local subsystem is based on the reduced unknown input observer (R-UIO) by [32]. Here, we exploit this R-UIO for the estimation of the physical coupling between local subsystems, by defining an extended system

$$\begin{cases} \bar{x}_i(k+1) = \bar{A}_{ii}\bar{x}_i(k) + \bar{B}_i u_i(k) + \bar{d}_i(k) \\ \quad y_i(k) = \bar{C}_i \bar{x}_i(k) \end{cases} \tag{12}$$

where $\bar{x}_i = \begin{bmatrix} x_i \\ d_i \end{bmatrix}$, $\bar{d}_i(k) = \begin{bmatrix} 0 \\ d_i(k+1) - d_i(k) \end{bmatrix}$, $\bar{A}_{ii} = \begin{bmatrix} A_{ii} & I \\ 0 & I \end{bmatrix}$, $\bar{C}_i = \begin{bmatrix} I & 0 \end{bmatrix}$, $\bar{B}_i = \begin{bmatrix} B_i \\ 0 \end{bmatrix}$. Then, the R-UIO takes the form

$$\begin{cases} \xi_i(k+1) = M_i \xi_i(k) + G_i u_i(k) + R_i y_i(k)\,, \\ \quad \hat{d}_i^{\ell}(k) = \xi_i(k) + H_i y_i(k)\,, \end{cases} \tag{13}$$

where ξ_i is the observer state, $\hat{d}_i^{\ell}$ is the local disturbance estimate, and M_i, G_i, R_i, and H_i are designed such that

$$\begin{aligned} ||M_i|| &< 1\,, \\ M_i T_i - R_i \bar{C}_i - T_i \bar{A}_i &= 0\,, \\ T_i + H_i \bar{C}_i - L_i &= 0\,, \\ G_i - T_i \bar{B}_i &= 0\,. \end{aligned} \tag{14}$$

An efficient design approach can be found in [32]. Following definition of the matrices in (14), the dynamics of the local estimation error $e_i^{\ell} = \hat{d}_i^{\ell} - d_i$ can be written as

$$e_i^{\ell}(k+1) = M_i e_i^{\ell}(k) - T_i \bar{d}_i(k)\,, \tag{15}$$

which, given (14), converges to a neighborhood of the origin asymptotically. Let us define $e^{\ell} = \mathrm{col}_{i\in\mathcal{N}}[e_i^{\ell}]$ such that the stability of (15) implies that $e^{\ell} \in \mathcal{E}^{\ell}$, where

$$\mathcal{E}^{\ell}(k+1) = M\mathcal{E}^{\ell}(k) \oplus (-T)\bar{\mathcal{D}}(k)\,, \tag{16}$$

where $M = \mathrm{diag}_{i\in\mathcal{N}}[M_i]$, $T = \mathrm{diag}_{i\in\mathcal{N}}[T_i]$ and $\bar{\mathcal{D}}(k)$ is defined as $\bar{\mathcal{D}}(k) = \mathcal{D}^s(k) \oplus (-\mathcal{D}^s(k+1))$. Note that (16) is evaluated at the supervisory level to obtain $\mathcal{E}^{\ell}$. Thus, because $\mathcal{D}^s$ is known, it can be used to obtain $\bar{\mathcal{D}}$. Furthermore, it can be seen the supervisory level requires no knowledge of the local dynamics to obtain $\mathcal{E}^{\ell}$, although it requires knowledge of the R-UIO parameters M and T.

Note that $\hat{d}_i^{\ell}$ is calculated at the local level using (13), while detection is performed at the supervisory level. As such, it must be transmitted from the local to the supervisory level. Therefore, given that the cyber-attack can fully compromise a local controller, the estimate $\hat{d}_i^{\ell}$ transmitted to the supervisory level might also be subject to a cyber-attack. To allow for this additional attack, we define the local estimate sent to the supervisory level as $\hat{d}_i^{\ell,a} = \hat{d}_i^{\ell} + a_i^d$. Based on (13) and (16), and the additional cyber-attack we define the *local estimation set* at the supervisory level as

$$\mathcal{D}^{\ell} = \hat{d}^{\ell,a} \oplus \mathcal{E}^{\ell}\,, \tag{17}$$

where $\hat{d}^{\ell,a} = \mathrm{col}_{i\in\mathcal{N}} \hat{d}_i^{\ell,a}$.

C. Cyber-Attack Detection Condition

At the supervisory level cyber-attack detection will be performed using sets $\mathcal{D}^{\ell}$ and $\mathcal{D}^s$ as previously defined. By construction, for $k < k_a$, the physical coupling satisfies

$$d \in \mathcal{D}^{\ell} \cap \mathcal{D}^s \triangleq \mathcal{D}\,. \tag{18}$$

As d itself is not known, the condition $d \notin \mathcal{D}$ cannot directly be used for cyber-attack detection. Alternatively, we can check the detection condition

$$\mathcal{D} = \emptyset\,, \tag{19}$$

which implies $d \notin \mathcal{D}$. A summary of the scheme is given in Algorithms 1 and 2.

Remark 3: All sets described in this section are defined as constrained zonotopes, which are closed under matrix multiplication, addition and intersection, thus justifying the equalities in (8), (16), and (18). ◁

Algorithm 1 Actions at Each Local Subsystem i

Initialize: Determine $\hat{d}_i^{\ell}(0)$.
  for all $k \geq 0$ **do**
    Compute $u_i(k)$ (3) and $\hat{d}_i^{\ell}(k)$ (13).
    Transmit $\hat{d}_i^{\ell,a}$ to the supervisory level.
  end for

Algorithm 2 Actions at the Supervisory Level

Initialize: Determine $\mathcal{E}^s(0)$ and $\mathcal{E}^{\ell}(0)$.
  for all $k \geq 0$ **do**
    for all $j \in \mathcal{N}$ **do**
      Determine r_j,
      Compute $\mathcal{E}_j$ (9, 10).
      Enclose $\mathcal{E}_j$ by $\bar{\mathcal{E}}_j$.
    end for
    Receive $\hat{d}^{\ell,a}$ from local subsystems.
    Compute $\hat{d}^s$ (7), $\mathcal{D}^s$ (8, 11), $\mathcal{D}^{\ell}$ (16, 17), $\mathcal{D}$ (18).
    Check the Detection Condition (19).
  end for

V. Robustness and Detectability Analysis

The properties of the proposed cyber-attack detection scheme are analysed based on changes in the local disturbance estimate $\hat{d}_i^{\ell,a}$, as it is the only part of the detection algorithm affected by the considered cyber-attacks. Let us introduce $\hat{d}_i^{\ell,0}$ and $\hat{d}^{\ell,0} = \mathrm{col}_{i\in\mathcal{N}} \hat{d}_i^{\ell,0}$ as the nominal local

estimate of the physical coupling, i.e. $\hat{d}^{\ell,0}(k) = \hat{d}^{\ell,a}(k)$ if $a_i^u(k) = a_i^d(k) = 0 \ \forall \ k, i \in \mathcal{N}$. Accordingly, we define $\mathcal{D}^{\ell,0} \triangleq \hat{d}^{\ell,0} \oplus \mathcal{E}^\ell$. Similarly, we define x_i^a as the part of the state driven by an attack a_i^u, satisfying dynamics: $x_i^a(k+1) = A_{ii}^{cl} x_i^a(k) + B_i a_i^u(k)$. Because of superposition in linear systems, x_i can be written as $x_i = x_i^0 + x_i^a$, where x_i^0 is the nominal state.

Theorem 1: It can be guaranteed no false detection occurs, i.e. $\mathcal{D}^s \cap \mathcal{D}^\ell \neq \emptyset$ for all $k \leq k_a$. □

Proof: The sets $\mathcal{D}^s$ and $\mathcal{D}^\ell$ are such that, in nominal conditions, $d \in \mathcal{D}^s$ and $d \in \mathcal{D}^\ell$ hold by construction. Therefore, $d \in \mathcal{D}^s \cap \mathcal{D}^\ell \neq \emptyset$ for all $k \leq k_a$ ■

Theorem 2: Any attack for which $\hat{d}^{\ell,a} - \hat{d}^{\ell,0} \in \mathcal{D}^{s'} \ominus \mathcal{D}^{\ell,0}$ is guaranteed to be detected. □

Proof: Define $\mathcal{A} = \mathcal{D}^{s'} \ominus \mathcal{D}^{\ell,0}$, such that $\mathcal{D}^{\ell,0} \oplus \mathcal{A} = \mathcal{D}^{s'}$, which implies $(\mathcal{D}^{\ell,0} \oplus \mathcal{A}) \cap \mathcal{D}^s = \emptyset$ [29]. This proves that an attack is detected if $\hat{d}^{\ell,a} - \hat{d}^{\ell,0} \in \mathcal{A}$. ■

To analyze existence conditions for stealthy attacks, let us introduce the following definitions.

Definition 1 (Locally Stealthy): A cyber-attack $\{a_i^u \neq 0, a_i^d \neq 0\}$ is locally stealthy if $\hat{d}_i^{\ell,a}(k) = \hat{d}_i^{\ell,0}(k) \ \forall k \geq 0$. ◁

Definition 2 (Globally Stealthy): A cyber-attack $\{a_i^u \neq 0, a_i^d \neq 0\}$ is globally stealthy if it is locally stealthy and $\hat{d}_j^{\ell,a}(k) = \hat{d}_j^{\ell,0}(k) \ \forall k, j$ s.t. $i \in \mathcal{N}_j$. ◁

Theorem 3: For all $i \in \mathcal{I}_a$, there exists $a_i^u \neq 0$ and $a_i^d \neq 0$ for which $\hat{d}_i^{\ell,a} = \hat{d}_i^{\ell,0}$, i.e. the attack is locally stealthy. □

Proof: For sake of space, the proofs of Theorem 3 and Theorem 4 are omitted. ■

Theorem 4: There exists a globally stealthy attack if and only if a_i^u and a_i^d satisfy Theorem 3, and a_i^u is such that $x_i^a \in \ker A_{ji}$, for all $j \in \{l \in \mathcal{N} : i \in \mathcal{N}_l\} \triangleq \widehat{\mathcal{N}}_i$, $i \in \mathcal{I}_a$. □

Remark 4: Although the construction of locally and globally stealthy attacks is not given, we note that to perform such attacks, malicious agents require information relating to more than only the attacked subsystems. Indeed, to be locally stealthy, a_i^d must be designed to consider the effect that x_i^a has on the neighbors of $i \in \mathcal{I}_a$ through the physical coupling. This is different to other definitions of locally stealthy attacks available in literature [21], where information about the local subsystem dynamics is sufficient to perform such attacks. ◁

VI. Simulation Example

To demonstrate the effectiveness of the proposed hierarchical detection scheme, we apply it to a system with four subsystems that are physically connected in series. The system is discretized with a time-step of 0.25 $[s]$ and modelled by (1) and (3) using the following parameters $\forall i \in \mathcal{N}$

$$A_{ii} = \begin{bmatrix} 1 & 0.25 \\ -0.055 & 0.995 \end{bmatrix}; B_i = \begin{bmatrix} 0 \\ 0.05 \end{bmatrix} \ ; \ K_i = \begin{bmatrix} -0.284 \\ -2.100 \end{bmatrix}^\top$$

$$L_i = 1.484 \ ; \ A_{ij} = \begin{bmatrix} 0.005 & 0 \\ 0 & 0 \end{bmatrix} \forall j \in \{i-1, i+1\} \cap \mathcal{N}$$

Cyber-attack detection is performed using Algorithms 1 and 2. Here the parameters for R-UIO (13) are found using [32]. Furthermore, b_i in (10) is chosen as 0.955. The references r_i that are non-zero are shown in Figure 2. The system is

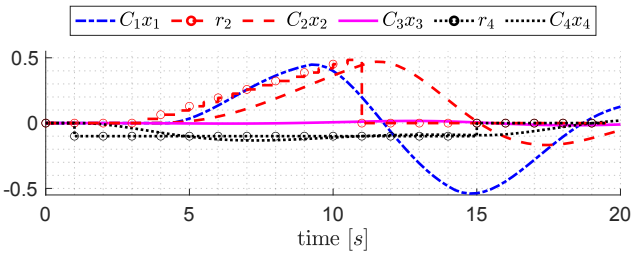

Fig. 2. Reference and Tracking performance of all subsystems.

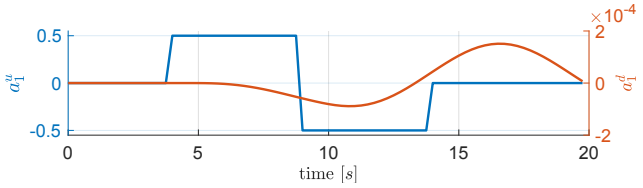

Fig. 3. The implemented attack, where a_1^d makes the attack locally stealthy.

corrupted by an attacker in subsystem 1 as shown in Figure 3. Here a_i^u disturbs the system and a_i^d is designed to make the attack locally stealthy (see Theorem 3).

Figure 2 also shows the reference tracking performance of all subsystems. One can see that all systems show reasonable tracking except the attacked subsystem 1. The detection of attack a_1^u is shown as red shade in Figure 4. Additionally, a metric is introduced which gives more insight into how close a detection is. This metric is denoted i_f and is defined as

$$i_f = \frac{\text{Vol}(\mathcal{D})}{\min(\text{Vol}(\mathcal{D}^s), \text{Vol}(\mathcal{D}^\ell))},$$

such that $i_f = 1$ if set $\mathcal{D}^\ell$ encloses set $\mathcal{D}^s$ or vice versa, and $i_f = 0$ if the sets do not intersect.

Figure 4 shows an estimate of i_f obtained using a Monte Carlo method using 1000 random samples of $\mathcal{D}^\ell$ and $\mathcal{D}^s$. This estimate is used as obtaining the exact volume of a constrained zonotope is computationally hard. One can see that i_f is normally 1 and becomes smaller than 1 around times of detection of attack a_1^u and momentarily around 1 and 3 $[s]$ due to excitation of the system. The temporary lack of detection at $12 - 14$ $[s]$ is caused by the change of sign of the attack at $8[s]$, which causes the system to temporarily be in a condition that resembles nominal behaviour.

Figure 5 shows a projection of the sets $\mathcal{D}^\ell$ and $\mathcal{D}^s$ on the axes representing d_2 and d_4. Due to the attack local estimation set $\mathcal{D}^\ell$ moves while the supervisory estimation set $\mathcal{D}^s$ remains unaffected. The projections shown in Figure 5 are used here to illustrate the detection method, and any use of these projections for identification has not been studied.

VII. Conclusion

Hierarchical control architectures are commonly used in industrial control systems, however very little work exists on fault or cyber-attack detection schemes with the same architecture. In this paper a hierarchical cyber-attack detection scheme has been presented which utilizes two estimates of the physical coupling between local subsystem. These are

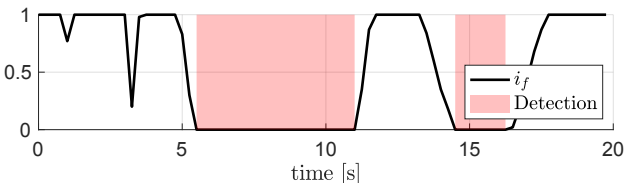

Fig. 4. Detection times and fraction of sets $\mathcal{D}^h$ and $\mathcal{D}^\ell$ that are intersecting.

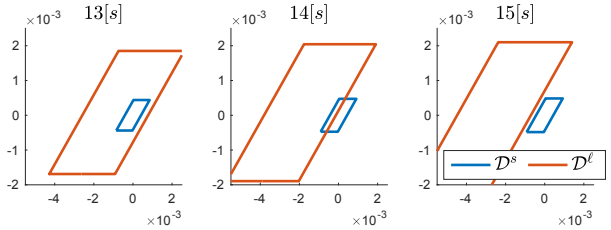

Fig. 5. Visualization of the set based attack detection using a projection to show the estimate sets for $\mathcal{S}_i, i \in \{2, 4\}$.

compared at the supervisory level to detect cyber-attacks on the local subsystems. The redundancy of information contained in these two estimates can be used for detection. Indeed, the local estimate uses local model knowledge and measurements to estimate how each subsystem is affected by the coupling; on the other hand, the supervisory estimate uses knowledge of the reference tracked by each subsystem to estimate how each subsystem affects its neighbours. A set based method using constrained zonotope representation is presented to calculate the estimation errors, which are compared for cyber-attack detection. It is proven that the detection method is robust, and existence conditions for guaranteed detectability, and locally and globally stealthy attacks are given. In future work, we intend to extend the proposed detection scheme to multi-rate systems.

REFERENCES

[1] R. Baheti and H. Gill, "Cyber-physical systems," *The impact of control technology*, vol. 12, no. 1, pp. 161–166, 2011.

[2] E. A. Lee, "Cyber physical systems: Design challenges," in *ISORC 2008*, 2008, pp. 363–369.

[3] N. Falliere, L. O. Murchu, and E. Chien, "W32. stuxnet dossier," *White paper, Symantec Corp., Security Response*, vol. 5, no. 6, p. 29, 2011.

[4] J. Giraldo, E. Sarkar, A. A. Cardenas, 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, 2017.

[5] H. Sandberg, S. Amin, and K. H. Johansson, "Cyberphysical security in networked control systems: An introduction to the issue," *IEEE Control Systems*, vol. 35, no. 1, pp. 20–23, 2015.

[6] Z. Chen, F. Pasqualetti, J. He, P. Cheng, H. L. Trentelman, and F. Bullo, "Guest editorial: Special issue on security and privacy of distributed algorithms and network systems," *IEEE Trans. on Autom. Control*, vol. 65, no. 9, pp. 3725–3727, 2020.

[7] J. Lunze, *Feedback control of large-scale systems*. Prentice Hall, 1992.

[8] D. D. Siljak, *Decentralized control of complex systems*. Courier Corporation, 2011.

[9] R. Scattolini, "Architectures for distributed and hierarchical model predictive control–a review," *J. process control*, vol. 19, no. 5, pp. 723–731, 2009.

[10] P. Chanfreut, J. M. Maestre, and E. F. Camacho, "A survey on clustering methods for distributed and networked control systems," *Annual Reviews in Control*, vol. 52, pp. 75–90, 2021.

[11] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, "Distributed fault diagnosis and fault-tolerant control," in *Diagnosis and Fault-Tolerant Control*. Springer, 2016, pp. 467–518.

[12] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Distributed fault detection and isolation resilient to network model uncertainties," *IEEE Trans. on Cybernetics*, vol. 44, no. 11, pp. 2024–2037, 2014.

[13] F. Arrichiello, A. Marino, and F. Pierri, "Observer-based decentralized fault detection and isolation strategy for networked multirobot systems," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 4, pp. 1465–1476, 2015.

[14] 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, 2016.

[15] F. Boem, R. M. G. Ferrari, C. Keliris, T. Parisini, and M. M. Polycarpou, "A distributed networked approach for fault detection of large-scale systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 18–33, 2017.

[16] F. Boem, A. J. Gallo, D. M. Raimondo, and T. Parisini, "Distributed fault-tolerant control of large-scale systems: An active fault diagnosis approach," *IEEE Transactions on Control of Network Systems*, vol. 7, no. 1, pp. 288–301, 2020.

[17] M. Khalili, X. Zhang, Y. Cao, M. M. Polycarpou, and T. Parisini, "Distributed fault-tolerant control of multiagent systems: An adaptive learning approach," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 2, pp. 420–432, 2020.

[18] M. Deghat, V. Ugrinovskii, I. Shames, and C. Langbort, "Detection and mitigation of biasing attacks on distributed estimation networks," *Automatica*, vol. 99, pp. 369–381, 2019.

[19] T. Keijzer, F. Jarmolowitz, and R. M. G. Ferrari, "Detection of cyber-attacks in collaborative intersection control," *ECC*, pp. 62–67, 2021.

[20] R. Anguluri, V. Katewa, and F. Pasqualetti, "Centralized versus decentralized detection of attacks in stochastic interconnected systems," *IEEE Trans. on Autom. Control*, vol. 65, no. 9, pp. 3903–3910, 2019.

[21] A. Barboni, H. Rezaee, F. Boem, and T. Parisini, "Detection of covert cyber-attacks in interconnected systems: A distributed model-based approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3728–3741, 2020.

[22] 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 Transactions on Automatic Control*, vol. 65, no. 9, pp. 3800–3815, 2020.

[23] S. A. Salinas and P. Li, "Privacy-preserving energy theft detection in microgrids: A state estimation approach," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 883–894, 2016.

[24] R. Nikoukhah, "guaranteed active failure detection and isolation for linear dynamical system," *Automatica*, vol. 34, no. 11, pp. 1345–1358, 1998.

[25] J. K. Scott, R. Findeisen, R. D. Braatz, and D. M. Raimondo, "Input design for guaranteed fault diagnosis using zonotopes," *Automatica*, vol. 50, no. 6, pp. 1580–1589, 2014.

[26] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz, "Constrained zonotopes: A new tool for set-based estimation and fault detection," *Automatica*, vol. 69, pp. 126–136, 2016.

[27] V. Puig, "Fault diagnosis and fault tolerant control using set-membership approaches: Application to real case studies," *International Journal of Applied Mathematics and Computer Science*, 2010.

[28] V. Rostampour, R. M. Ferrari, A. M. Teixeira, and T. Keviczky, "Privatized distributed anomaly detection for large-scale nonlinear uncertain systems," *IEEE Transactions on Automatic Control*, vol. 66, no. 11, pp. 5299–5313, 2020.

[29] F. Blanchini and S. Miani, *Set-theoretic methods in control*. Springer, 2008, vol. 78.

[30] B. A. Francis, "The linear multivariable regulator problem," *SIAM J. on Control and Optimization*, vol. 15, no. 3, pp. 486–505, 1977.

[31] D. Barcelli, A. Bemporad, and G. Ripaccioli, "Decentralized hierarchical multi-rate control of constrained linear systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 277–283, 2011.

[32] J. Lan and R. J. Patton, "A new strategy for integration of fault estimation within fault-tolerant control," *Automatica*, vol. 69, pp. 48–59, 2016.