Distributed Fault Detection and Isolation of Large-scale Discrete-time Nonlinear Systems: an Adaptive Approximation Approach

Riccardo M. G. Ferrari, Thomas Parisini, and Marios M. Polycarpou

Abstract—This paper deals with the problem of designing a distributed fault detection and isolation methodology for nonlinear uncertain large-scale discrete-time dynamical systems. As a *divide et impera* approach is used to overcome the scalability issues of a centralized implementation, the large scale system being monitored is modelled as the interconnection of several subsystems. The subsystems are allowed to overlap, thus sharing some state components. For each subsystem, a *Local Fault Diagnoser* is designed, based on the measured local state of the subsystem as well as the transmitted variables of neighboring states that define the subsystem interconnections. The local diagnostic decision is made on the basis of the knowledge of the local subsystem dynamic model and of an adaptive approximation of the interconnection with neighboring subsystems. The use of a specially-designed consensus-based estimator is proposed in order to improve the detectability and isolability of faults affecting variables shared among overlapping subsystems. Theoretical results are provided to characterize the detection and isolation capabilities of the proposed distributed scheme. Finally, simulation results are reported showing the effectiveness of the proposed distributed fault diagnosis methodology.

I. INTRODUCTION

The problem of automated fault diagnosis and accommodation is motivated by the need to develop more autonomous and intelligent systems that operate reliably in the presence of system faults. In dynamical systems, faults are characterized by critical and unpredictable changes in the system dynamics, thus requiring the design of suitable fault diagnosis schemes [1], [2], [3]. Moreover, with current technological trends several systems of practical interest are large-scale and/or physically distributed and thus the decomposition and spatial distribution of highly demanding computational tasks is of critical importance.

Recently there has been significant research activity in modeling, control and cooperation methodologies for distributed systems (see, for example, [4], and the references cited therein). This activity is motivated by several applications, especially in complex large-scale systems, such as traffic networks, environmental systems, communication networks, power grid networks, water distribution networks, etc. Such systems, although their dynamics and control objectives may appear to be completely different, have some important common characteristics: their dynamics are complex and spatially distributed, and, as a result, it is typically more convenient to decompose the system into smaller subsystems which can be more easily controlled and monitored locally (or regionally). The study of controlling spatially distributed systems is not a

Corresponding author: Thomas Parisini. This work has been partially supported by the Italian Ministry for University and Research, by Regione Friuli-Venezia-Giulia, and the Research Promotion Foundation of Cyprus.

Riccardo M.G. Ferrari is with Danieli Automation S.p.A., Buttrio, Italy. (r.ferrari@dca.it)

T. Parisini is with Imperial College London (UK) and University of Trieste (I). (t.parisini@paperplaza.net)

M. M. Polycarpou is with the KIOS Research Center for Intelligent Systems and Networks, Dept. of Electrical and Computer Engineering, University of Cyprus, Nicosia 1678, Cyprus. (mpolycar@ucy.ac.cy)

new problem. As far back as in the 1970s, researchers sought to develop so called "decentralized control" methods [5]. Since then there have been many advancements in the design and analysis of distributed control schemes. On the other hand, much less research activity has been devoted at the problem of designing fault diagnosis schemes specifically for distributed systems.

Due to the complexity of the problem, in practice it is difficult to achieve robust fault diagnosis in large-scale distributed systems within a centralized implementation, mainly because of scalability issues. In fact, a centralized scheme sooner or later may hit one of the two following constraints on the hardware/software architecture used to implement it: limited available computation power for evaluating the fault decision, and limited communication bandwidth for acquiring all the necessary measurements. While considerable effort was aimed at developing distributed fault diagnosis algorithms suited to discrete event systems (see, for instance, [6]), much less attention was devoted to discrete or continuous–time systems (see [7], where the problem of designing sensor networks for fault-tolerant estimation is addressed, [8], [9] where fault-tolerance in distributed systems is considered, [10], [11], [12], which are focused on decentralized fault detection, and [13] dealing with fault consensus in networks of unmanned vehicles).

In previous works [14], [15], [16], the authors developed some preliminary results on a quantitative distributed fault detection scheme where a large-scale system was decomposed into a set of disjoint subsystems, and the physical interaction between neighboring subsystems was described by uncertain nonlinear functions. A network of *Local Fault Detectors* (LFDE) was developed so that each LFDE monitored a single subsystem by making use of the measurement of local variables, as well as the value of some *interconnection variables* communicated by neighboring LFDs. But, apart from this exchange of measurements, the neighboring LFDEs were not involved in the process of deciding whether a fault occurred in a subsystem. In this paper, the above distributed detection scheme is extended to allow cooperation between neighboring LFDEs by using *overlapping decompositions* [17] of the initial large-scale system. In this way, more than one LFDE may be monitoring a single *shared variable* and collectively decide on the presence of faults influencing it. This is implemented by means of a specially designed consensus-like estimation scheme that may improve the capability of the LFDEs to detect a fault with respect to the consensus–less, non overlapping case.

In [16], a distributed fault detection scheme for continuous–time systems based on overlapping decompositions of subsystems, in turn monitored by a network of interconnected local fault detectors has been proposed. The novel contribution of the present paper, in the context of discrete-time nonlinear systems, is the extension of this idea by including also local and global fault isolation capabilities, thanks to the introduction of specialized *Fault Isolation Estimators* and a *Global*

Fault Diagnoser (see [18] for some preliminary results). A rigorous characterization of the fault isolation capability of the proposed scheme is given, while an 11–tanks system is used throughout the paper to illustrate the decomposition strategy, the modeling of local and distributed faults and, finally, to show the effectiveness of the proposed methodology by extensive simulation results.

The paper is organized as follows: in Section II, a problem formulation is developed for fault diagnosis of distributed dynamical systems. The design and analysis of a distributed fault detection and isolation architecture is presented in Section III, followed by the detailed development of its detection part in Section IV, and of its isolation part in Section V. Finally, simulation results for illustrating the methodology are given in Section VI, while Section VII provides some concluding remarks.

II. BACKGROUND

Let us consider a nonlinear dynamic system $\mathscr{S}$, referred to as *monolithic system* and described by the following discrete–time model

$$\mathscr{S} : \ \boldsymbol{x}(t+1) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)) + \boldsymbol{\eta}(t) + \beta(t-T_0)\boldsymbol{\phi}(\boldsymbol{x}(t), \boldsymbol{u}(t)) , \quad (1)$$

where $t \in \mathbb{N}$ is the discrete-time instant, $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ and $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{m}}$ denote[1] the state and input vectors, respectively, and $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{n}} \times \mathbb{R}^{\boldsymbol{m}} \mapsto \mathbb{R}^{\boldsymbol{n}}$ represents the *nominal healthy dynamics*. Moreover, the function $\boldsymbol{\eta} : \mathbb{N} \mapsto \mathbb{R}^{\boldsymbol{n}}$ stands for the uncertainty in the state equation and includes external disturbances as well as modeling errors and possibly discretization errors. From a qualitative viewpoint, the term $\beta(t-T_0)\boldsymbol{\phi}(\boldsymbol{x}(t), \boldsymbol{u}(t))$ represents the deviation in the system dynamics due to a fault. The term $\beta(t - T_0)$ characterizes the time profile of a fault that occurs at some *unknown* discrete-time instant T_0, and $\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{u})$ denotes the nonlinear fault function. This formulation (first introduced in [19]) allows both additive and multiplicative faults (since $\boldsymbol{\phi}$ is a function of $\boldsymbol{x}$ and $\boldsymbol{u}$), as well as more general nonlinear faults. The fault time profile $\beta(t - T_0)$ models *incipient* faults characterized by an exponential decaying time-profile

$$\beta(t - T_0) = \begin{cases} 0 & \text{if} \quad t < T_0 , \\ 1 - b^{-(t-T_0)} & \text{if} \quad t \geq T_0 \end{cases}, \quad (2)$$

where $b \geq 1$ denotes the unknown fault-evolution rate (the case of an "abrupt" fault time-profile can be obtained as $b \to \infty$ in (2)). Note that the fault time profile given by (2) only reflects the developing rate of the fault, while all its other basic features are captured by the function $\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{u})$, which describes the changes in the dynamics due to the fault.

The problem of detecting and isolating faults in nonlinear uncertain systems described by (1) using adaptive approximation methodologies has been addressed in several works in the literature (see, among others, [20], [21] and the references cited therein). In this paper, we consider design and analysis of an adaptive approximation methodology for the case of large-scale and distributed nonlinear systems for which a centralized *Fault Detection and Isolation* (FDI) architecture may not be possible or not desirable. As the impossibility usually derives from a centralized implementation being unable to process and/or convey all the necessary measurements at a single computation node in real time, in this paper a *divide et impera* approach is used, in order to decompose the (possible large) original FDI problem into a number of smaller problems, easier to solve with the available hardware/software infrastructure.

Therefore, we consider the decomposition of system $\mathscr{S}$ into N subsystems[2] $\mathscr{S}_I$, $I = 1, \ldots, N$, each characterized by a *local* state vector $x_I \in \mathbb{R}^{n_I}$, with a separate monitoring *agent* designed for each $\mathscr{S}_I$.

In order to introduce system decompositions, first of all the system *structure* is defined using graph theory [22].

Definition 1: The *structure* $\Sigma_{\mathscr{S}}$ of a dynamical system $\mathscr{S}$ having a state vector $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ and an input vector $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{m}}$ is the set of ordered pairs

$$\Sigma_{\mathscr{S}} \triangleq \{(\boldsymbol{x}^{(\boldsymbol{i})}, \boldsymbol{x}^{(\boldsymbol{j})}) : \boldsymbol{i}, \boldsymbol{j} \in \{\boldsymbol{1}, \ldots, \boldsymbol{n}\}, "\boldsymbol{x}^{(\boldsymbol{i})} \text{ affects } \boldsymbol{x}^{(\boldsymbol{j})}"\}$$
$$\cup \ \{(\boldsymbol{u}^{(\boldsymbol{i})}, \boldsymbol{x}^{(\boldsymbol{j})}) : \boldsymbol{i} \in \{\boldsymbol{1}, \ldots, \boldsymbol{m}\}, \boldsymbol{j} \in \{\boldsymbol{1}, \ldots, \boldsymbol{n}\},$$
$$"\boldsymbol{u}^{(\boldsymbol{i})} \text{ affects } \boldsymbol{x}^{(\boldsymbol{j})}"\} .$$

Definition 2: The *structural graph* [17] of a dynamical system $\mathscr{S}$, having a state vector $\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}$ and an input vector $\boldsymbol{u} \in \mathbb{R}^{\boldsymbol{m}}$, is the directed graph (*digraph*) $\mathcal{G} \triangleq \{\mathcal{N}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}}\}$ having the node set $\mathcal{N}_{\mathcal{G}} \triangleq \{\boldsymbol{x}^{(\boldsymbol{i})} : \boldsymbol{i} \in \{\boldsymbol{1}, \ldots, \boldsymbol{n}\}\} \cup \{\boldsymbol{u}^{(\boldsymbol{i})} : \boldsymbol{i} \in \{\boldsymbol{1}, \ldots, \boldsymbol{m}\}\}$ and the system structure $\Sigma_{\mathscr{S}}$ as the arc set, that is $\mathcal{E}_{\mathcal{G}} = \Sigma_{\mathscr{S}}$.

The decomposition of the monolithic system $\mathscr{S}$ is based on decomposing its structural graph. The idea of graph decomposition has been used in many fields [23]. For example, graph decomposition has been used in numerical methods involving the solution to partial differential equations [24], [25], [26], [27], in image processing [28], in operations research [29], and, of course, in large–scale system decomposition [17], [30].

To decompose a monolithic system $\mathscr{S}$ described as in (1) and having a structural graph $\mathcal{G} = (\mathcal{N}_{\mathcal{G}}, \mathcal{E}_{\mathcal{G}})$, we define $N \geq 1$ subsystems $\mathscr{S}_I$, with $I \in \{1, \ldots, N\}$, each one having a *local state* vector $x_I \in \mathbb{R}^{n_I}$ and a *local input* vector $u_I \in \mathbb{R}^{m_I}$. These local vectors are constructed by taking components of the monolithic system vectors $\boldsymbol{x}$ and $\boldsymbol{u}$, based on ordered sets $\mathcal{I}_I \triangleq (\mathcal{I}_I^{(1)}, \ldots, \mathcal{I}_I^{(n_I)})$ of indices, called *extraction index set* [15], [31], [16]. These sets can be defined by introducing the following *extraction mapping* between local and monolithic indexes.

Definition 3: For each subsystem $\mathscr{S}_I$, its extraction index set $\mathcal{I}_I$ is obtained by means of an *extraction mapping* $\sigma_I : \{1, \ldots, n_I\} \mapsto \{\boldsymbol{1}, \ldots, \boldsymbol{n}\}$, so that $\mathcal{I}_I \triangleq (\sigma_I(1), \ldots, \sigma_I(n_I))$.

Definition 4: The *local state* $x_I \in \mathbb{R}^{n_I}$ and the *local input* $u_I \in \mathbb{R}^{m_I}$ of a dynamical subsystem $\mathscr{S}_I$, arising from the decomposition of a monolithic system $\mathscr{S}$, are respectively the vectors $x_I \triangleq \mathrm{col}(\boldsymbol{x}^{(\boldsymbol{j})} : \boldsymbol{j} \in \mathcal{I}_I)$ and $u_I \triangleq \mathrm{col}(\boldsymbol{u}^{(\boldsymbol{k})} : (\boldsymbol{u}^{(\boldsymbol{k})}, \boldsymbol{x}^{(\boldsymbol{j})}) \in \mathcal{E}_{\mathcal{G}}, \boldsymbol{j} \in \mathcal{I}_I, \boldsymbol{k} = \boldsymbol{1}, \ldots, \boldsymbol{m})$, where $\mathcal{I}_I$ is the extraction index set of the I–th subsystem.

It is worth noting that, when performing the "col" operation in the two previous definitions, the elements of the index set $\mathcal{I}_I$ are taken in the order they appear. According to Definition 4, the local input contains all the input components that affect at least one component of the local state vector. At this point, the structural graph of the I–th subsystem can be easily defined as the subgraph $\mathcal{G}_I$ induced on $\mathcal{G}$ by the subset made of all the components of x_I together with those of u_I. The following provides a definition for the decomposition of a large-scale system.

Definition 5: A *decomposition* $\mathcal{D}$ of dimension N of the large-scale system $\mathscr{S}$ is a multiset $\mathcal{D} \triangleq \{\mathscr{S}_1, ..., \mathscr{S}_N\}$ made of N subsystems, defined through a multiset $\{\mathcal{I}_1, \ldots, \mathcal{I}_N\}$ of index sets, such that for each $I \in \{1, \ldots, N\}$ the following axioms hold:

1) $\mathcal{I}_I \neq \emptyset$;
2) $\mathcal{I}_I^{(j)} \leq \boldsymbol{n}$, for each $j \in \{1, \ldots, n_I\}$;

---

[1] Here and in the rest of the paper the use of bold letters indicates that a given quantity is related to the monolithic system.

[2] In the paper, a capital-case index denotes a specific sub-system.

3) the subdigraph of $\mathcal{G}$ induced by $\mathcal{I}_I$ must be weakly connected, that is, each component of x_I must act on or must be acted on by at least another component of x_I;

4) $\bigcup_{I=1}^{N} \mathcal{I}_I = \{1, ..., n\}$.

Axiom 1 prevents the definition of trivial empty subsystems, Axiom 2 is necessary for well-posedness, Axiom 3 avoids that resulting subsystems have isolated state components, while Axiom 4 requires that the decomposition covers the whole original monolithic system. It is important to note that the above decomposition does not require that for any two subsystems $\mathcal{I}_I \cap \mathcal{I}_J = \emptyset$, $I, J \in \{1, \ldots, N\}$. This allows for a state component of $\mathcal{S}$ to be assigned to more than one subsystems, thus being "shared". Such a decomposition is called *overlapping decomposition*. Overlapping decompositions [32] have been found to be useful tools when addressing large–scale systems. In particular, problems of stability, control and estimation [33], and fault diagnosis [34] for large–scale linear systems were addressed using overlapping decompositions.

As a result of overlaps, some components of x are assigned to more than one subsystems thus giving rise to the concepts of *shared state variable* and *overlap index set*.

Definition 6: A shared state variable $x^{(s)}$ is a component of x such that $s \in \mathcal{I}_I \cap \mathcal{I}_J$, for some $I, J \in \{1, \ldots N\}$, $I \neq J$ and a given decomposition $\mathcal{D}$ of dimension N.

Definition 7: The *overlap index set* of subsystems sharing a variable $x^{(s)}$ is the set $\mathcal{O}_s \triangleq \{I : s \in \mathcal{I}_I\}$, whose cardinality is N_s.

In the following, the notation $x_I^{(s_I)}$, with $x_I^{(s_I)} \equiv x^{(s)}$, will be used to denote the fact that the s–th state component of the original large–scale system, after the decomposition became the s_I–th of the I–th subsystem, $I \in \mathcal{O}_s$.

Now, we define the interaction (if any) between different subsystems. In this framework, the external variables influencing the dynamics of local state components of subsystem $\mathcal{S}_I$ making up the vector of *interconnection variables* z_I.

Definition 8: The *interconnection variables* vector $z_I \in \mathbb{R}^{p_I}$, $(p_I \leq n - n_I)$ of the subsystem $\mathcal{S}_I$ is the vector $z_I \triangleq \mathrm{col}(x^{(k)} : (x^{(k)}, x^{(j)}) \in \mathcal{E}_{\mathcal{G}}, j \in \mathcal{I}_I, k \in \{1, \ldots, n\})$.

The set of subsystems acting on a given subsystem $\mathcal{S}_I$ through the interconnection vector z_I is the *neighbors index set* $\mathcal{J}_I$, a concept that naturally leads to the introduction of the *fundamental graph* [17] whose nodes represent subsystems and whose arcs represent their interaction through interconnection variables.

Definition 9: The *neighbors index set* of a subsystem $\mathcal{S}_I$ is the set $\mathcal{J}_I \triangleq \{K : \exists (x^{(k)}, x^{(j)}) \in \mathcal{E}_{\mathcal{G}}, k \in \mathcal{I}_K, j \in \mathcal{I}_I, K \in \{1, \ldots, N\} \setminus \{I\}\}$.

Definition 10: The *fundamental graph* of a distributed system, obtained by applying the decomposition $\mathcal{D}$ to the monolithic system $\mathcal{S}$, is the digraph $\mathcal{G}_{\mathcal{D}} \triangleq \{\mathcal{N}_{\mathcal{G}_{\mathcal{D}}}, \mathcal{E}_{\mathcal{G}_{\mathcal{D}}}\}$, where the node set $\mathcal{N}_{\mathcal{G}_{\mathcal{D}}} \triangleq \{\mathcal{S}_I : I \in \{1, \ldots, N\}\}$ contains the subsystems and the arc set is $\mathcal{E}_{\mathcal{G}_{\mathcal{D}}} \triangleq \{(\mathcal{S}_J, \mathcal{S}_I) : I \in \{1, \ldots, N\}, J \in \mathcal{J}_I\}$.

Unlike linear systems, for which powerful model decomposition techniques and descriptions exist (see for instance the works published in recent years by D'Andrea et al. [35], [4]), that can be applied to systems showing either a regular or arbitrary structure, for nonlinear systems the decomposition task is much more difficult, and in general it is not possible to devise an additive decomposition into purely local and purely interconnection terms. Therefore, a general decomposition as in [17] is considered:

$$\mathcal{S}_I : x_I(t+1) = f_I(x_I(t), u_I(t)) + g_I(x_I(t), z_I(t), u_I(t)) \\ + \beta(t - T_0)\phi_I(x_I(t), z_I(t), u_I(t)), \quad (3)$$

where $f_I : \mathbb{R}^{n_I} \times \mathbb{R}^{m_I} \mapsto \mathbb{R}^{n_I}$ is the *local nominal* function and $g_I : \mathbb{R}^{n_I} \times \mathbb{R}^{p_I} \times \mathbb{R}^{m_I} \mapsto \mathbb{R}^{n_I}$ represents the *interconnection function*, where the effects of the local uncertainty term η_I has also been incorporated, with $\eta_I \triangleq \mathrm{col}(\eta^{(j)} : j \in \mathcal{I}_I)$. Furthermore, $u_I \in \mathbb{R}^{m_I}$, $(m_I \leq m)$ is the *local input* (see Definition 4), $z_I \in \mathbb{R}^{p_I}$, $(p_I \leq n - n_I)$ is the vector of *interconnection variables* (see Definition 8), and $\phi_I : \mathbb{R}^{n_I} \times \mathbb{R}^{p_I} \times \mathbb{R}^{m_I} \mapsto \mathbb{R}^{n_I}$ is the *local fault function*.

Let us now introduce two assumptions that will be used in the subsequent analysis.

Assumption 1: The fault function ϕ is such that the fundamental graph $\mathcal{G}_{\mathcal{D}}$ remains the same before and after the fault event.

Assumption 1 is introduced to simplify the formal analysis; according to this assumption, in the paper we suppose that the possible fault event does not cause a change to the system structure $\Sigma_{\mathcal{S}}$ by adding new dependencies between variables belonging to different subsystems, so that it is possible to write the local fault function ϕ_I as a function of local and interconnection variables only[3]. This also means that the neighbors index set $\mathcal{J}_I$ and the interconnection variables vector z_I do not change structure due to the occurrence of a fault.

Assumption 2: For each $\mathcal{S}_I, I = 1, \ldots, N$, the state variables $x_I(t)$ and control variables $u_I(t)$ remain bounded before and after the occurrence of a fault, i.e., there exist some stability regions $\mathcal{R}_I = \mathcal{R}_I^x \times \mathcal{R}_I^u \subset \mathbb{R}^{n_I} \times \mathbb{R}^{m_I}$, such that $(x_I(t), u_I(t)) \in \mathcal{R}_I^x \times \mathcal{R}_I^u$, $\forall I = 1, \ldots, N, \forall t \geq 0$. Finally, the time profile parameter b is unknown but it is lower bounded by a known constant $\bar{b}$, that is $0 < \bar{b} \leq b$.

As a consequence of Assumption 2, for each subsystem $\mathcal{S}_I, I = 1, \ldots, N$, it is possible to define some stability regions $\mathcal{R}_I^z$ for the interconnecting variables z_I. Since no fault accommodation is considered in this work (only fault detection and isolation), the feedback controller acting on the system $\mathcal{S}$ must be such that the variables $x(t)$ and $u(t)$ remain bounded for all $t \geq 0$. However, it is important to state in advance that the design of the distributed FDI methodology does not depend on the specific structure of the underlying controller and hence the controller design is not discussed. Assumption 2 is required for well-posedness, but does not cause major loss of generality to the proposed FDI scheme. In fact, from a practical perspective, detecting faulty modes characterized by large or even unbounded "magnitudes" typically turns out to be quite an easy task by resorting to limit-checking techniques.

The interconnection function g_I in the decomposition described by (3) includes the uncertainty represented by the term η_I. Therefore, in the sequel the following further assumption will be needed.

Assumption 3: The interconnection function g_I is an unstructured and uncertain nonlinear function, whose k–th component is bounded by some known function[4], i.e.,

$$|g_I^{(k)}(x_I, z_I, u_I)| \leq \bar{g}_I^{(k)}(x_I, z_I, u_I), \quad \forall x_I \in \mathcal{R}_I^x, \\ \forall z_I \in \mathcal{R}_I^z, \forall u_I \in \mathcal{R}_I^u, \quad (4)$$

where the bounding function $\bar{g}_I^{(k)} \geq 0$ is known and bounded for all $I = 1, \ldots, N$.

A. Example

To gain some more insight into the afore-described decomposition approach, consider the example depicted in Fig. 1,

---

[3]However, it is possible for a fault event to remove some of the interconnections, which can be formally represented by setting some g_I function to zero.

[4]In the paper, when there is no risk of ambiguity and for the sake of simplicity, a compact notation like, for instance, $g_I(t) \equiv g_I(x_I(t), z_I(t), u_I(t))$, is used.

where a specific decomposition of a system $\mathscr{S}$ into three overlapping subsystems $\mathscr{S}_1$, $\mathscr{S}_2$ and $\mathscr{S}_3$ is considered. The example of Fig. 1 deals with a 11–tank system, which will be re-considered in the simulation Section VI.
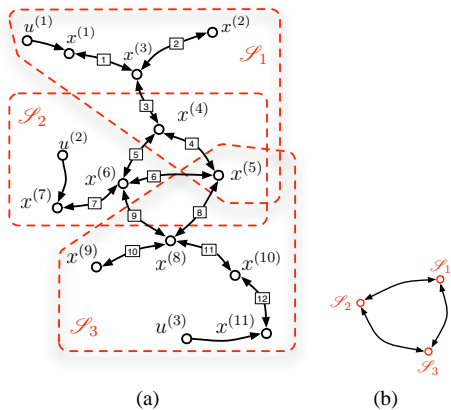

Figure 1. (a) Example of decomposition of a system $\mathscr{S}$ into three overlapping subsystems $\mathscr{S}_1$, $\mathscr{S}_2$ and $\mathscr{S}_3$, and (b) the corresponding fundamental graph.

The decomposition shown in this example is such that: $x_1 = [\boldsymbol{x^{(1)}}, \boldsymbol{x^{(2)}}, \boldsymbol{x^{(3)}}, \boldsymbol{x^{(4)}}, \boldsymbol{x^{(5)}}]^\top$, $x_2 = [\boldsymbol{x^{(4)}}, \boldsymbol{x^{(5)}}, \boldsymbol{x^{(6)}}, \boldsymbol{x^{(7)}}]^\top$ and $x_3 = [\boldsymbol{x^{(5)}}, \boldsymbol{x^{(8)}}, \boldsymbol{x^{(9)}}, \boldsymbol{x^{(10)}}, \boldsymbol{x^{(11)}}]^\top$ are the local states, $u_1 = \boldsymbol{u^{(1)}}$, $u_2 = \boldsymbol{u^{(2)}}$ and $u_3 = \boldsymbol{u^{(3)}}$ are the local inputs, $z_1 = [\boldsymbol{x^{(6)}}, \boldsymbol{x^{(8)}}]^\top$, $z_2 = [\boldsymbol{x^{(3)}}, \boldsymbol{x^{(8)}}]^\top$ and $z_3 = [\boldsymbol{x^{(4)}}, \boldsymbol{x^{(6)}}]^\top$ are the interconnection variables. Furthermore, $\boldsymbol{x^{(4)}} \equiv x_1^{(4)} \equiv x_2^{(1)}$ and $\boldsymbol{x^{(5)}} \equiv x_1^{(5)} \equiv x_2^{(2)} \equiv x_3^{(1)}$ are shared variables with $\mathcal{O}_4 = \{1,2\}$ and $\mathcal{O}_5 = \{1,2,3\}$.

III. DISTRIBUTED FAULT DETECTION AND ISOLATION ARCHITECTURE

The backbone of the proposed *Distributed Fault Detection and Identification* (DFDI) architecture is made of N communicating agents called *Local Fault Diagnosers* (LFDs) $\mathscr{L}_I$, which are devoted to monitor each of the N subsystems. The LFDs generate a *fault decision* d_I^{FD} regarding the mode of behavior (healthy or one among the possible faulty modes) of the corresponding subsystem $\mathscr{S}_I$. These decisions are gathered by a higher level agent $\mathscr{L}$, which is referred to as *Global Fault Diagnoser* (GFD), in order to coordinate the LFDs and formulate a fault decision d^{FD} about the health of the global system $\mathscr{S}$. Fig. 2 shows in pictorial form the structure of the DFDI architecture using the same illustrative example of a distributed system already presented in Fig. 1. The various part of the architecture are arranged in three layers: the first layer is constituted of the physical subsystems, the second layer is made up by local fault diagnosers, while the third one contains the global fault diagnoser. The different type of arrows highlight the different interactions between the parts of the architecture: physical interactions in the first layer, consistent information flows between layer one and two and between parts of layer two, while sporadic communication between the second and the third layer is illustrated by dashed arrows. More details on the structure of the LFDs and of the GFD will be provided later on.

Following the fault isolation formulation proposed in [21], for isolation purposes we assume that for the global system $\mathscr{S}$ there exists a *global fault set* $\mathcal{F}$ containing $N_{\mathcal{F}}$ possible nonlinear fault functions $\boldsymbol{\phi_l}(\boldsymbol{x}, \boldsymbol{u})$, $\boldsymbol{l} \in \{1, \ldots, N_{\mathcal{F}}\}$. Following
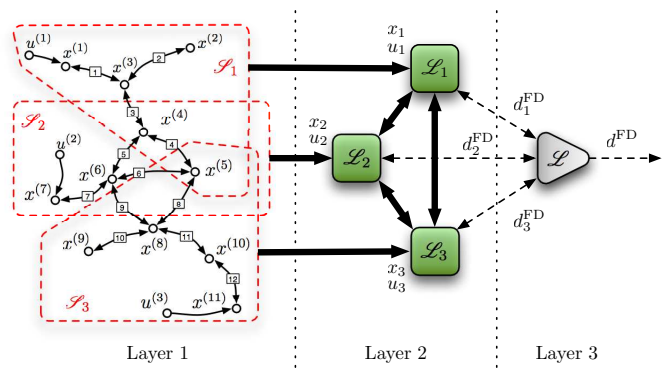

Figure 2. A scheme of the proposed DFDI architecture concerning the same three subsystems of Fig. 1 , $\mathscr{S}_1$,

the decomposition $\mathcal{D}$ and based on Assumption 1, the introduction of the global fault set $\mathcal{F}$ leads to the existence, for each subsystem $\mathscr{S}_I$, of a *local fault set* $\mathcal{F}_I$ containing $N_{\mathcal{F}_I}$ known types of possible nonlinear fault functions[5] $\phi_{I,l}(x_I, z_I, u_I)$, $l \in \{1, \ldots, N_{\mathcal{F}_I}\}$. Thus, each LFD $\mathscr{L}_I$ provides a fault decision d_I^{FD} regarding the health of the corresponding subsystem $\mathscr{S}_I$, by relying on $N_{\mathcal{F}_I}+1$ nonlinear adaptive estimators of the local state x_I, with $I \in \{1, \ldots, N\}$. The first estimator, called *Fault Detection Approximation Estimator* (FDAE), is based on the nominal model (3) and is used for fault detection. The remaining $N_{\mathcal{F}_I}$ estimators, called *Fault Isolation Estimators* (FIE), make up a bank of estimators to be used to determine which of the possible $N_{\mathcal{F}_I}$ faults in the set $\mathcal{F}_I$ has occurred.

Under normal operating conditions (that is from $t = 0$ until a fault is detected) the FDAE is the only estimator that each LFD employs. After a fault is detected by *any* of the N LFDs, the GFD receives the corresponding local fault decision, and in response triggers the switch of each LFD from fault detection to fault isolation operating mode. In the latter mode, each LFD activates its own bank of FIEs in order to try to locally isolate the occurred fault, by employing kind of a *Generalized Observer Scheme* (GOS), (see [36], [37]). The local fault decisions d_I^{FD} of the LFDs are communicated to the GFD, allowing it to determine which one of the faults in the global set $\mathcal{F}$, if any, affects the system $\mathscr{S}$ (see Section V and Algorithm 1).

In the DFDI scheme, we assume that every LFD takes uncertain measurements of x_I according to $y_I(t) \triangleq x_I(t) + \xi_I(t)$, where ξ_I is an unknown term characterizing the measurement error associated with the process of measuring x_I by each LFD (we assume u_I to be perfectly available). Moreover, each LFD communicates with the neighboring LFDs in $\mathcal{J}_I$ in order to fill the interconnection vector z_I (see the example in Fig. 2). Due to the uncertain state measurements, it follows that, instead of receiving the actual interconnection vector z_I, each LFD receives from its neighbors the vector $v_I(t) \triangleq z_I(t) + \zeta_I(t)$, where $\zeta_I(t)$ is made of the components of ξ_J affecting the relevant components of the measurements y_J, $J \in \mathcal{J}_I$.

Assumption 4: The measuring uncertainties represented by the vectors ξ_I and ζ_I are unstructured and unknown, but, for each $h = 1, \ldots, n_I$ and for each $k = 1, \ldots, p_I$, the components of ξ_I and of ζ_I are bounded, respectively, as

$$|\xi_I^{(h)}(t)| \le \bar{\xi}_I^{(h)}, \quad |\zeta_I^{(k)}(t)| \le \bar{\zeta}_I^{(k)}, \quad \forall t \ge 0, \qquad (5)$$

where $\bar{\xi}_I^{(h)}$ and $\bar{\zeta}_I^{(k)}$ are known positive scalars. Hence, it is possible to define *a priori* two compact regions of interest $\mathcal{R}_I^\xi$

[5]The global and the local fault sets is described in detail in Sec. V.

and $\mathcal{R}_I^\zeta$ such that $\xi_I(t) \in \mathcal{R}_I^\xi$ and $\zeta_I(t) \in \mathcal{R}_I^\zeta$.

Under the assumptions made so far, a shared variable $\boldsymbol{x}^{(\boldsymbol{s})}$ is measured by distinct LFDs in the overlap set $\mathcal{O}_{\boldsymbol{s}}$ with distinct uncertainties. Furthermore, because of Assumption 3, the interconnection part of the local model (3) may also be affected by distinct uncertainties. Following these considerations, in the sequel, a cooperation mechanism between LFDs in the overlap set $\mathcal{O}_{\boldsymbol{s}}$ will be devised in order to improve the overall diagnosis performances by exploiting the distributed nature of the fault diagnosis technique.

IV. DISTRIBUTED FAULT DETECTION

After the DFDI algorithm is initialized at $t = 0$ by turning on each I-th LFD, only its FDAE estimator is enabled and monitors the subsystem $\mathscr{S}_I$, providing a *local state estimate* $\hat{x}_{I,0}$ of the local state x_I. The difference between the estimate $\hat{x}_{I,0}(t)$ and the measurements $y_I(t)$ yields the *estimation error* $\epsilon_{I,0}(t) \triangleq y_I(t) - \hat{x}_{I,0}(t)$ which plays the role of a residual and will be compared, component by component, with a suitable *detection threshold*[6] $\bar{\epsilon}_{I,0}(t) \in \mathbb{R}_+^{n_I}$. The following condition

$$|\epsilon_{I,0}^{(k)}(t)| \le \bar{\epsilon}_{I,0}^{(k)}(t), \quad \forall\, k = 1, \ldots, n_I \qquad (6)$$

is associated with the *fault-free hypothesis*

$$\mathscr{H}_{I,0} : \text{"The system } \mathscr{S}_I \text{ is healthy"}. \qquad (7)$$

By this, we mean that (6) is a necessary (but generally not sufficient) condition for (7), so that should condition (6) be violated at some time instant t, then the hypothesis $\mathscr{H}_{I,0}$ is falsified and the so–called *local fault detection signature* $\mathcal{S}_{I,0}$ is generated, thus leading to a local fault detection decision. In qualitative fault diagnosis schemes, such as [38], the fault signature is defined as a symbolic vector, that qualitatively describes the behavior of residuals and their derivatives after the occurrence of a fault. Instead, in quantitative schemes, such as [1], [2], [36], the fault signature represents the pattern of residuals that exhibit abnormal behavior after the occurrence of a fault. In this regard, we introduce a few further useful definitions.

Definition 11: The *local detection signature* associated with the subsystem $\mathscr{S}_I$, $I \in \{1, \ldots, N\}$ at the discrete-time instant $t > 0$ is the index set

$$\mathcal{S}_{I,0}(t) \triangleq \{ k \in \{1, \ldots, n_I\} : \exists\, t_1,\, t \ge t_1 > 0$$
$$\text{such that } |\epsilon_{I,0}^{(k)}(t_1)| > \bar{\epsilon}_{I,0}^{(k)}(t_1) \}. \quad (8)$$

In relation with the *fundamental graph* introduced in Definition 10, the fundamental detection signature can be defined as follows.

Definition 12: The *fundamental detection signature* associated with the system $\mathscr{S}$ at the discrete-time instant $t > 0$ is the index set

$$\mathcal{S}(t) \triangleq \{ I \in \{1, \ldots, N\} : \mathcal{S}_{I,0}(t) \ne \emptyset \}. \qquad (9)$$

Now, the local fault detection logic for the I–th LFD can be stated in terms of the local detection signature $\mathcal{S}_{I,0}(t)$. Specifically, a fault affecting the I–th subsystem is detected by its LFD at the first discrete-time instant $\bar{t}$ such that $\mathcal{S}_{I,0}(\bar{t})$ becomes non-empty. This discrete-time instant is called the *local fault detection time* $T_{I,d}$, as formally defined in the following.

Definition 13: The *local fault detection time* $T_{I,d}$ is defined as $T_{I,d} \triangleq \min\{t : \mathcal{S}_{I,0}(t) \ne \emptyset\}$.

[6]To be defined in eq. (20).

Finally, the *fault detection time* T_d is simply defined as the earliest among the local detection times.

Definition 14: The *fault detection time* T_d is defined as $T_d \triangleq \min\{t : \mathcal{S}(t) \ne \emptyset\}$.

This formalizes the fact that in the proposed architecture the event of a LFD detecting a fault is immediately relayed to the global fault diagnoser $\mathscr{L}$. The GFD computes the fundamental detection signature $\mathcal{S}$ and sets T_d as the earliest discrete-time instant at which it becomes non empty. Then, it immediately informs every LFD that a fault has been detected in the system and that the isolation mode, introduced in Section III and further described in Section V, should be activated.

Remark 1: The communication between the LFDs and the GFD required to implement the DFDI architecture is event-driven, that is, only events such as the detection or isolation of a fault are communicated through the channels depicted as dashed arrows in Fig. 2. As this kind of exchanged information is limited to simple boolean values, this means that even if the communication between the LFDs and the GFD follows a one-to-all pattern, scalability should not be an issue in practical applications.

A. Local Fault Detection and Approximation Estimator

The local FDAE is a nonlinear adaptive estimator based on the subsystem model (3), which (as in [16] in the continuos-time case) generalizes to the distributed context the fault diagnosis methodology presented in [21].

First of all, the simpler case of a non-shared state variable is addressed. The estimate of the k-th component $\hat{x}_{I,0}^{(k)}$ is computed as

$$\hat{x}_{I,0}^{(k)}(t+1) = \lambda(\hat{x}_{I,0}^{(k)}(t) - y_I^{(k)}(t)) + f_I^{(k)}(y_I(t), u_I(t))$$
$$+ \hat{g}_I^{(k)}(y_I(t), v_I(t), u_I(t), \hat{\vartheta}_{I,0}), \quad (10)$$

where $0 < \lambda < 1$. Following the idea presented in [16], the term $\hat{g}_I^{(k)}$ is the k–th output of an adaptive approximator designed to learn the unknown interconnection function g_I, and $\hat{\vartheta}_{I,0} \in \hat{\Theta}_{I,0}$ denotes its adjustable parameters vector, with $\hat{\Theta}_{I,0} \subset \mathbb{R}^{q_{I,0}}$ being a compact set[7]. As in [16], in this paper we assume that $\hat{g}_I^{(k)}$ represents a linear-in-the-parameters, but otherwise nonlinear multivariable approximation model, such as neural networks, fuzzy logic networks, polynomials, spline functions, wavelet networks, etc.

It is important to emphasize the differences among the present approach and the one described in [21] regarding the centralized case. Whilst in [21] the adaptive approximator is devoted to learn the fault function after the detection of a fault, in the present case, the adaptive approximator starts from the very beginning to learn the uncertain interconnection function in order to facilitate more accurate and faster detection. It is worth noting that to implement (10), the I-th LFD needs only to receive from its neighbors the values of the variables making up the the interconnection vector $v_I(t)$.

In order for $\hat{g}_I$ to learn the interconnection function g_I, the parameter vector $\hat{\vartheta}_{I,0}$ is updated according to the following law:

$$\hat{\vartheta}_{I,0}(t+1) = \mathcal{P}_{\hat{\Theta}_{I,0}} \left[\hat{\vartheta}_{I,0}(t) + \gamma_{I,0}(t) H_{I,0}^\top(t) r_{I,0}(t+1) \right],$$

where $H_{I,0}(t) \triangleq \partial \hat{g}_I(t)/\partial \hat{\vartheta}_{I,0} \in \mathbb{R}^{n_I \times q_{I,0}}$ denotes the gradient matrix of the on–line approximator with respect

[7]For the sake of simplicity we assume $\hat{\Theta}_{I,0}$ to be a origin–centered hypersphere with radius $M_{\hat{\Theta}_{I,0}}$ (see [21] for some remarks on this geometrical simplification).

to its adjustable parameters, and $r_{I,0}(t + 1)$ is given by $r_{I,0}(t + 1) = \epsilon_{I,0}(t + 1) - \lambda\epsilon_{I,0}(t)$. $\mathcal{P}_{\hat{\Theta}_{I,0}}$ is a projection operator [39] restricting $\hat{\vartheta}_{I,0}$ within $\hat{\Theta}_{I,0}$ according to:

$$\mathcal{P}_{\hat{\Theta}_{I,0}}(\hat{\vartheta}_{I,0}) \triangleq \begin{cases} \hat{\vartheta}_{I,0} & \text{if } |\hat{\vartheta}_{I,0}| \leq M_{\hat{\Theta}_{I,0}} \\ M_{\hat{\Theta}_{I,0}}\hat{\vartheta}_{I,0}/|\hat{\vartheta}_{I,0}| & \text{if } |\hat{\vartheta}_{I,0}| > M_{\hat{\Theta}_{I,0}} \end{cases}.$$

The learning rate $\gamma_{I,0}(t)$ is computed at each step t as $\gamma_{I,0}(t) \triangleq \mu_{I,0}/(\varepsilon_{I,0} + \|H_{I,0}^\top(t)\|_F^2)$, with $\varepsilon_{I,0} > 0$, $0 < \mu_{I,0} < 2$, where $\|\cdot\|_F$ denotes the Frobenius norm and $\varepsilon_{I,0}, \mu_{I,0}$ are design constants that guarantee the stability of the learning law [39], [40].

The case of a variable $x^{(s)}$ of the original centralized system $\mathscr{S}$ that, after the decomposition, is shared among more than one LFDs (see the simple example shown in Fig. 1) is more complicated. Clearly, one option is for each LFD to just implement its own version of the recurrent equation (10), by using the measurement $y_J^{(s_J)}$, the local model $f_J^{(s_J)}$ and the components $\hat{g}_J^{(s_J)}$ of the adaptive interconnection approximator. Instead, in order to take advantage of the redundancy introduced by the overlap and motivated by the encouraging practical results shown in [16], in this paper we use a deterministic consensus scheme between the LFDs in $\mathcal{O}_s$ so that their FDAEs cooperate towards the estimation of the shared state variable $x^{(s)}$. The proposed consensus protocol leads to the following FDAE dynamic equation for the generic I–th LFD, $I \in \mathcal{O}_s$:

$$\hat{x}_{I,0}^{(s_I)}(t + 1) = \lambda(\hat{x}_{I,0}^{(s_I)}(t) - y_I^{(s_I)}(t))$$
$$+ W_s^{(I,I)}[f_I^{(s_I)}(y_I(t), u_I(t)) + \hat{g}_I^{(s_I)}(y_I(t), v_I(t), u_I(t), \hat{\vartheta}_{I,0})]$$
$$+ \lambda \sum_{J \in \mathcal{O}_s \setminus \{I\}} W_s^{(I,J)} \left[\hat{x}_{J,0}^{(s_J)}(t) - \hat{x}_{I,0}^{(s_I)}(t)\right]$$
$$+ \sum_{J \in \mathcal{O}_s \setminus \{I\}} W_s^{(I,J)}[f_J^{(s_J)}(y_J(t), u_J(t))$$
$$+ \hat{g}_J^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{J,0})], \quad (11)$$

where the additional terms with respect to (10) appearing in the second line smooth out the difference between the various estimate of the shared variable, and those in the third line average the various local functions and approximated interconnection functions. It is of customary importance to note that, in order to implement (11), the LFD $\mathscr{L}_I$ does not need the information about the expressions of $f_J^{(s_J)}$ and of $\hat{g}_J^{(s_J)}$; instead, it suffices that $\mathscr{L}_J$, $J \in \mathcal{O}_s$, computes *locally* the term $f_J^{(s_J)} + \hat{g}_J^{(s_J)}$ and communicates it to the other LFDs, according to a suitable communication graph $\mathcal{G}_s$, alongside its actual state estimate $\hat{x}_{J,0}^{(s_J)}$. Specifically, for the sake of generality, we assume to have a generic communication graph $\mathcal{G}_s \triangleq (\mathcal{O}_s, \mathcal{E}_s)$, that may include the all-to-all communication as a special case. Bearing this in mind, (11) can be rewritten in more compact form as

$$\hat{x}_{I,0}^{(s_I)}(t + 1) = \lambda\{\hat{x}_{I,0}^{(s_I)}(t) - y_I^{(s_I)}(t)$$
$$+ \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[\hat{x}_{J,0}^{(s_J)}(t) - \hat{x}_{I,0}^{(s_I)}(t)]\}$$
$$+ \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[f_J^{(s_J)}(y_J(t), u_J(t))$$
$$+ \hat{g}_J^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{J,0})]. \quad (12)$$

The term $W_s = [W_s^{(I,J)}]$ is a weighted adjacency matrix reflecting the way the various LFDs estimating the same shared variable $x^{(s)}$ communicate with each other. In this work, only doubly-stochastic adjacency matrices $W_s \in \mathbb{R}^{N_s \times N_s}$ are considered [41]. For example, we may consider the *Metropolis* adjacency matrices [42], [43] defined as

$$W_s^{(I,J)} \triangleq \begin{cases} 0 & \text{, if } (I, J) \notin \mathcal{E}_s \\ 1/(1 + \max\{d_s^{(I)}, d_s^{(J)}\}) & \text{, if } (I, J) \in \mathcal{E}_s, I \neq J \\ 1 - \sum_{K \neq I} W_s^{(I,K)} & \text{, if } I = J \end{cases}$$
$$(13)$$

where $d_s^{(I)}$ is the degree of the I–th node in $\mathcal{G}_s$.

Remark 2: Requiring the matrix W_s to be doubly stochastic is a standard assumption in many problems of distributed control and estimation. As previously said, there exist simple weights selection schemes such as the Metropolis or the *Maximum-degree* [42] that guarantee double-stochasticity. Further details on existence conditions for doubly stochastic adjacency matrices, which arguments are based on the Birkhoff-von Neumann theorem, can be found in [44].

Before the occurrence of a fault (i.e., for $t < T_0$), the dynamics of the LFD estimation error component $\epsilon_{I,0}^{(s_I)}$ can be written as

$$\epsilon_{I,0}^{(s_I)}(t + 1) = \lambda\{\epsilon_{I,0}^{(s_I)}(t)$$
$$+ \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[\epsilon_{J,0}^{(s_J)}(t) - \epsilon_{I,0}^{(s_I)}(t) + \xi_I^{(s_I)}(t) - \xi_J^{(s_J)}(t)]\}$$
$$+ \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[f_J^{(s_J)}(x_J(t), u_J(t)) - f_J^{(s_J)}(y_J(t), u_J(t))$$
$$g_J^{(s_J)}(t) - \hat{g}_J^{(s_J)}(t)] + \xi_I^{(s_I)}(t + 1).$$

Since $\sum_{I \neq J} W_s^{(I,J)} = 1 - W_s^{(I,I)}$ by assumption, the estimation error dynamics satisfies

$$\epsilon_{I,0}^{(s_I)}(t+1) = \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}\{\lambda[\epsilon_{J,0}^{(s_J)}(t) - \xi_J^{(s_J)}(t)] + \Delta f_J^{(s_J)}(t)$$
$$+ \Delta g_J^{(s_J)}(t)\} + \lambda\xi_I^{(s_I)}(t) + \xi_I^{(s_I)}(t+1), \quad (14)$$

where the following scalar quantities are defined $\Delta f_J^{(s_J)}(t) \triangleq f_J^{(s_J)}(x_J(t), u_J(t)) - f_J^{(s_J)}(y_J(t), u_J(t))$, $\Delta g_J^{(s_J)}(t) \triangleq g_J^{(s_J)}(x_J(t), z_J(t), u_J(t)) - \hat{g}_J^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{I,0})$. Accordingly, the vectors Δf_I and Δg_I are defined as $\Delta f_I(t) \triangleq f_I(x_I(t), u_I(t)) - f_I(y_I(t), u_I(t))$ and $\Delta g_I(t) \triangleq g_I(x_I(t), z_I(t), u_I(t)) - \hat{g}_I(y_I(t), v_I(t), u_I(t), \hat{\vartheta}_{I,0})$.

It is worth noting that, in general, the functions $\Delta f_I(t)$ and $\Delta g_I(t)$ take on non-zero values due to several factors, including measurement errors on x_I, the measurement errors of neighbouring LFDs, and the uncertainty in the interconnection function g_I itself. Although the aim of the adaptive approximator $\hat{g}_I$ is to learn the uncertain function g_I, generally it cannot be expected to match the actual term g_I even if the weights of the adaptive approximator could be optimally selected. This may be formalized by introducing an *optimal weight vector* $\hat{\vartheta}_{I,0}^*$ [45]

$$\hat{\vartheta}_{I,0}^* \triangleq \arg \min_{\hat{\vartheta}_{I,0} \in \Theta_{I,0}} \sup_{x_I, z_I, u_I} \|g_I(x_I, z_I, u_I)$$
$$- \hat{g}_I(x_I, z_I, u_I, \hat{\vartheta}_{I,0})\|,$$

with x_I, z_I, u_I taking values in their respective domains. This leads to the definition of the *Minimum Functional Approximation Error* (MFAE) ν_I, which describes the least possible

approximation error that can be achieved at the discrete-time instant t if $\hat{\vartheta}_{I,0} = \hat{\vartheta}_{I,0}^*$:

$$\nu_I(t) \triangleq g_I(x_I(t), z_I(t), u_I(t)) - \hat{g}_I(x_I(t), z_I(t), u_I(t), \hat{\vartheta}_{I,0}^*).$$

By introducing the *parameter estimation error* $\tilde{\vartheta}_{I,0} \triangleq \hat{\vartheta}_{I,0}^* - \hat{\vartheta}_{I,0}$ and the following function

$$\Delta \hat{g}_I(t) \triangleq \hat{g}_I(x_I(t), z_I(t), u_I(t), \hat{\vartheta}_{I,0}) - \hat{g}_I(y_I(t), v_I(t), u_I(t), \hat{\vartheta}_{I,0}),$$

it turns out that $\Delta g_I(t)$ can be written as $\Delta g_I(t) = H_{I,0}\tilde{\vartheta}_{I,0} + \nu_I(t) + \Delta \hat{g}_I(t)$.

By using (14), the dynamics of the LFD estimation error component $\epsilon_{I,0}^{(s_I)}$ before the occurrence of a fault (i.e., $t < T_0$) can be written as

$$\epsilon_{I,0}^{(s_I)}(t+1) = \sum_{J \in \mathcal{O}_s} W_s^{(I,J)} \left[\lambda \epsilon_{J,0}^{(s_J)}(t) + \chi_J^{(s_J)}(t) \right] + \lambda \xi_I^{(s_I)}(t) + \xi_I^{(s_I)}(t+1), \quad (15)$$

where we introduced the following *total uncertainty* term

$$\chi_I^{(s_I)}(t) \triangleq \Delta f_I^{(s_I)}(t) - \lambda \xi_I^{(s_I)}(t) + \Delta g_I^{(s_I)}(t).$$

In order to analyze the behavior of $\epsilon_{I,0}^{(s_I)}(t)$ and define the threshold $\bar{\epsilon}_{I,0}^{(s_I)}(t)$ (see (6)), it is convenient to introduce the following vectors related to the detection estimator of all the LFDs sharing the variable $x^{(s)}$: $\epsilon_{s,0}(t) \triangleq \text{col}(\epsilon_{I,0}^{(s_I)}, I \in \mathcal{O}_s)$, $\chi_s(t) \triangleq \text{col}(\chi_I^{(s_I)}, I \in \mathcal{O}_s)$, and $\xi_s(t) \triangleq \text{col}(\xi_I^{(s_I)}, I \in \mathcal{O}_s)$.
The FDAE estimation error dynamics of all the LFDs in $\mathcal{O}_s$ can then be written in a more useful and compact form:

$$\epsilon_{s,0}(t+1) = W_s[\lambda \epsilon_{s,0}(t) + \chi_s(t)] + \lambda \xi_s(t) + \xi_s(t+1). \quad (16)$$

Since $\lambda < 1$ and W_s is a doubly stochastic matrix, all its eigenvalues are within the unitary circle. Then, it follows that (16) represents the dynamics of a stable LTI discrete–time system. The solution of (16) is

$$\epsilon_{s,0}(t) = W_s\{\lambda[\sum_{h=0}^{t-2}(\lambda W_s)^{t-2-h}(W_s\chi_s(h) + \lambda\xi_s(h) + \xi_s(h+1)) + \lambda^{t-1}W_s^{t-1}\epsilon_{s,0}(0)] + \chi_s(t-1)\} + \lambda\xi_s(t-1) + \xi_s(t), \quad (17)$$

so that, component-wise, it becomes

$$\epsilon_{I,0}^{(s_I)}(t) \equiv \epsilon_{s,0}^{(I)}(t) = w_{s,I}^\top\{\lambda[\sum_{h=0}^{t-2}(\lambda W_s)^{t-2-h}(W_s\chi_s(h) + \lambda^{t-1}W_s^{t-1}\epsilon_{s,0}(0) + \lambda\xi_s(h) + \xi_s(h+1)) + \chi_s(t-1)\} + \lambda\xi_s(t-1) + \xi_s^{(I)}(t), \quad (18)$$

where $w_{s,I}^\top$ is a vector containing the I–th row of matrix W_s.
Now, a threshold on the estimation error that guarantees no false–positive fault detections for $t < T_0$ is proposed. The absolute value of the estimation error for $t < T_0$ can be upper bounded by using the triangular inequality as follows:

$$|\epsilon_{I,0}^{(s_I)}(t+1)| \leq \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[\lambda|\epsilon_{J,0}^{(s_J)}(t)| + \bar{\chi}_J^{(s_J)}(t)] + \lambda\bar{\xi}_I^{(s_I)}(t) + \bar{\xi}_I^{(s_I)}(t+1), \quad (19)$$

where (upper bound on the total uncertainty term)[8]

$$\bar{\chi}_J^{(s_J)}(t) \triangleq \max_{\xi_J}|\Delta f_J^{(s_J)}(t)| + \|H_{J,0}\|\kappa_{J,0}(\hat{\vartheta}_{J,0}) + \bar{\nu}_J(t) + \lambda\bar{\xi}_J^{(s_J)}(t) + \max_{\xi_J}\max_{\zeta_J}|\Delta\hat{g}_J(t)|,$$

with the function $\kappa_{J,0}$ being such that[9] $\kappa_{J,0}(\hat{\vartheta}_{J,0}) \geq \|\tilde{\vartheta}_{J,0}\|$.
By taking the absolute value component-wise so that $|\epsilon_{s,0}| \equiv \text{col}(|\epsilon_{I,0}^{(s_I)}| : I \in \mathcal{O}_s)$, the inequalities (19), can be written as

$$|\epsilon_{s,0}(t+1)| \leq W_s[\lambda|\epsilon_{s,0}(t)| + \bar{\chi}_s(t)] + \lambda\bar{\xi}_s(t) + \bar{\xi}_s(t+1).$$

Using the Comparison Lemma [46], the absolute value of each component of ϵ_s can be bounded by the corresponding component of $\bar{\epsilon}_s$, defined as the solution of the following equation

$$\bar{\epsilon}_s(t+1) = W_s[\lambda\bar{\epsilon}_s(t) + \bar{\chi}_s(t)] + \lambda\bar{\xi}_s(t) + \bar{\xi}_s(t+1), \quad (20)$$

with initial conditions $\bar{\epsilon}_s(0) \triangleq \text{col}(\bar{\epsilon}_I^{(s_I)}(0) : I \in \mathcal{O}_s)$. It is worth noting that the adaptive threshold defined in (20) can be easily implemented by any LFD in $\mathcal{O}_s$ by means of a linear discrete-time first-order filter driven by a suitable input (see [21] in the continuous-time case).

B. Faulty behavior and Fault Detectability

In this subsection, the behavior of the DFDI algorithm in the presence of a fault and its detection capabilities is investigated. Assume that at the discrete-time instant $t = T_0$ a fault ϕ occurs. Let

$$\phi_s(x, u) = \text{col}(\phi^{(s)}(x, u), s = 1, \dots, n) \quad (21)$$

with $\phi^{(s)}$ denoting the component of the fault function affecting the s-th state equation of the monolithic system (see (1)). After the occurrence of the fault, for $t \geq T_0$, the estimation error dynamics for a shared state variable $x^{(s)}$ given by (16) becomes

$$\epsilon_{s,0}(t+1) = W_s[\lambda\epsilon_{s,0}(t) + \chi_s(t)] + (1 - b^{-(t-T_0)})\phi_s(t) + \lambda\xi_s(t) + \xi_s(t+1), \quad (22)$$

where $\phi_s(t) \in \mathbb{R}^{N_s}$ is a vector whose components are all equal to $\phi^{(s)}$. The following theorem gives a sufficient condition for the estimation error to cross its corresponding threshold in finite time, thus allowing the fault to be detected by the I-th LFD. Therefore, it characterizes the class of faults that can be detected by the proposed scheme, given the bounds available on the unknown functions.
Theorem 1 (Local Fault Detectability): Given a subsystem $\mathscr{S}_I$, if there exists a discrete-time instant $t_1 > T_0$ such that the fault ϕ_I satisfies the inequality

$$\left| \sum_{h=T_0}^{t_1-1} \lambda^{t_1-1-h}(1 - b^{-(h-T_0)})\phi_I^{(s_I)}(h) \right| > 2\bar{\epsilon}_{I,0}^{(s_I)}(t_1), \quad (23)$$

for at least one component $s_I \in \{1, \dots, n_I\}$, then the fault is detected at the discrete-time instant t_1, that is $|\epsilon_{I,0}^{(s_I)}(t_1)| > \bar{\epsilon}_{I,0}^{(s_I)}(t_1)$.

---

[8]The notation $\max_{\xi_J}$ is short for $\max_{\xi_J \in \mathcal{R}^{\xi_J}}$.
[9]As $\Theta_{J,0}$ is a compact the function $\kappa_{J,0}$ can always be defined.

Proof: At the discrete-time instant $t_1 > T_0$, by using (17) and (21), the estimation error vector $\epsilon_{\boldsymbol{s},0}$ can be written as

$$\epsilon_{\boldsymbol{s},0}(t) = \sum_{h=0}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}[\,W_{\boldsymbol{s}}\chi_{\boldsymbol{s}}(h) + \lambda\xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1)$$
$$+ (1 - b^{-(h-T_0)})\phi_{\boldsymbol{s}}(h)\,] + \lambda^t (W_s)^t \epsilon_{s,0}(0)\,. \quad (24)$$

By applying the same expansion as in equations (17) and (18), the solution for the estimation error for the s_I–th component of the I–th subsystem can be written as[10]

$$\epsilon_{I,0}^{(s_I)}(t_1) = w_{\boldsymbol{s},I}^{\top}\{\lambda[\sum_{h=0}^{t_1-2}(\lambda W_{\boldsymbol{s}})^{t_1-2-h}(W_{\boldsymbol{s}}\chi_{\boldsymbol{s}}(h)$$
$$+ \lambda\xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1)) + \lambda^{t_1-1}W_{\boldsymbol{s}}^{t_1-1}\epsilon_{\boldsymbol{s},0}(0)] + \chi_{\boldsymbol{s}}(t_1-1)\}$$
$$+ \lambda\xi_{\boldsymbol{s}}^{(I)}(t_1-1) + \xi_{\boldsymbol{s}}^{(I)}(t_1) + \sum_{h=T_0}^{t_1-1}\lambda^{t_1-1-h}(1-b^{-(h-T_0)})\phi_I^{(s_I)}(h)\,.$$

Using the triangle inequality, we obtain

$$|\epsilon_{I,0}^{(s_I)}(t_1)| \geq -|w_{\boldsymbol{s},I}^{\top}\lambda\sum_{h=0}^{t_1-2}(\lambda W_{\boldsymbol{s}})^{t_1-2-h}(W_{\boldsymbol{s}}\chi_{\boldsymbol{s}}(h)+\lambda\xi_{\boldsymbol{s}}(h)$$
$$+ \xi_{\boldsymbol{s}}(h+1))| - |\lambda^{t_1}w_{\boldsymbol{s},I}^{\top}W_{\boldsymbol{s}}^{t_1-1}\epsilon_{\boldsymbol{s},0}(0)| - |w_{\boldsymbol{s},I}^{\top}\chi_{\boldsymbol{s}}(t_1-1)|$$
$$- |\lambda\xi_{\boldsymbol{s}}(t_1)| - |\xi_{\boldsymbol{s}}(t_1)| + |\sum_{h=T_0}^{t_1-1}\lambda^{t_1-1-h}(1-b^{-(h-T_0)})\phi_I^{(s_I)}(h)|\,.$$

The threshold can be written as

$$\bar{\epsilon}_{I,0}^{(s_I)}(t_1) = w_{\boldsymbol{s},I}^{\top}\{\lambda[\sum_{h=0}^{t_1-2}(\lambda W_{\boldsymbol{s}})^{t_1-2-h}(W_{\boldsymbol{s}}\bar{\chi}_{\boldsymbol{s}}(h) + \lambda\bar{\xi}_{\boldsymbol{s}}(h)$$
$$+ \bar{\xi}_{\boldsymbol{s}}(h+1)) + \lambda^{t_1-1}W_{\boldsymbol{s}}^{t_1-1}\bar{\epsilon}_{\boldsymbol{s},0}(0)] + \bar{\chi}_{\boldsymbol{s}}(t_1-1)\}$$
$$+ \lambda\bar{\xi}_{\boldsymbol{s}}^{(I)}(t_1-1) + \bar{\xi}_{\boldsymbol{s}}^{(I)}(t_1)\,.$$

Now, from the definition of the threshold $\bar{\epsilon}_{I,0}^{(s_I)}$ in Subsection IV-A, it follows that the last inequality is implied by

$$|\epsilon_{I,0}^{(s_I)}(t_1)| \geq -\bar{\epsilon}_{I,0}^{(s_I)}(t_1) + |\sum_{h=T_0}^{t_1-1}\lambda^{t_1-1-h}(1-b^{-(h-T_0)})\phi_I^{(s_I)}(h)|\,,$$

so that the fault detection condition $|\epsilon_{I,0}^{(s_I)}(t_1)| \geq \bar{\epsilon}_{I,0}^{(s_I)}(t_1)$ is implied by the theorem hypothesis. ∎

Remark 3: Theorem 1 provides a (possibly conservative) sufficient condition for fault detectability: if at some discrete-time instant $t > T_0$ at least one subsystem shows a non-empty local detection signature $\mathcal{S}_{I,0}(t)$, then this would cause the GFD to be alerted by the corresponding LFD. In qualitative and rough terms, the inequality on the left-hand side of (23) characterizes the relative "magnitude" of the effect of the fault versus the upper bound on the unknown functions quantified by the right-hand side of (23). It is also worth noting that be easily made specific to the case of non–shared variables.

---

[10]As $W_{\boldsymbol{s}}$ is doubly stochastic and all the components of $\phi_{\boldsymbol{s}}$ are equal to $\phi^{(s)}$, it holds $(W_s)^h\phi_{\boldsymbol{s}} = \phi_{\boldsymbol{s}}$ for all h.

V. DISTRIBUTED FAULT ISOLATION

A. Formulation of the distributed fault isolation problem

For isolation purposes, it is assumed that the fault function ϕ may either be unknown or belong to a known global fault set $\mathcal{F}$

$$\mathcal{F} \triangleq \{\boldsymbol{\phi_1}(\boldsymbol{x},\boldsymbol{u}), \ldots, \boldsymbol{\phi_{N_{\mathcal{F}}}}(\boldsymbol{x},\boldsymbol{u})\}\,.$$

In general, not all the subsystems are affected by a given fault function ϕ_l, but only those in the corresponding *fault influence set* $\mathcal{U}_l$. For each l–th fault, $\mathcal{U}_l$ contains the indexes of all the subsystems $\mathscr{S}_I$ that, after the decomposition $\mathcal{D}$, are assigned at least a global state component $\boldsymbol{x}^{(s)}$ for which the fault function ϕ_l is non–zero for at least one discrete-time instant, as defined below.

Definition 15: The *fault influence set* $\mathcal{U}_l$ for the l–th fault function ϕ_l is the index set

$$\mathcal{U}_l \triangleq \{I : \exists t, \exists s, \, s \in \mathcal{I}_I, \, \phi_l^{(s)}(\boldsymbol{x}(t), \boldsymbol{u}(t)) \neq 0\}\,. \quad (25)$$

For each subsystem $\mathscr{S}_I$, a *local fault set* $\mathcal{F}_I$ (defined below) can be built with the local fault functions obtained by all the global faults ϕ_l such that $I \in \mathcal{U}_l$:

$$\mathcal{F}_I \triangleq \{\phi_{I,1}(x_I, z_I, u_I), \ldots, \phi_{I,N_{\mathcal{F}_I}}(x_I, z_I, u_I)\}\,.$$

Notice that the local fault functions depend only on the local variables x_I, z_I and u_I (see Assumption (1)). The global index $\boldsymbol{l}$ and the local index l_I of a fault are related by a mapping $\varsigma_I : \{1, \ldots, N_{\mathcal{F}_I}\} \mapsto \{1, \ldots, N_{\mathcal{F}}\}$, so that $\boldsymbol{l} = \varsigma_I(l_I)$. This means that, for all the subsystems $\mathscr{S}_I$ so that $I \in \mathcal{U}_l$, for the generic component $\boldsymbol{s} \in \mathcal{I}_I$ of a global fault function it holds that $\boldsymbol{\phi_l^{(s)}}(\boldsymbol{x},\boldsymbol{u}) \equiv \phi_{I,l_I}^{(s_I)}(x_I, z_I, u_I)$, with $\boldsymbol{s} = \sigma_I(s_I)$.

The concept of the fault influence sets naturally leads to a subdivision of the faults into two categories, depending upon their topology: *local faults*, whose influence set is a singleton, and *distributed faults*, whose influence set includes more than one subsystem. Now, these categories are illustrated in the context of the same simple example of Fig. 1.

1) Local Fault: The simplest situation is exemplified in Figure 3a. The structure of the fault ϕ_1 is enhanced: dashed arcs represent part of the healthy dynamics changed by the fault, and filled nodes represent variables affected by the fault. As can be seen, the arc 1 is faulty so that the dynamics of the variables $\boldsymbol{x^{(1)}}$ and $\boldsymbol{x^{(3)}}$ are affected, thus leading to the fault influence set being $\mathcal{U}_1 = \{1\}$. This implies that only the local detection signature $\mathcal{S}_{1,0}$ (see (8)) may become non-empty as this fault affects only variables "internal" to subsystem $\mathscr{S}_1$ that are not shared by any other subsystems. More precisely, if the first LFD detects a fault at a discrete-time instant $T_{1,d}$, then the local detection signature satisfies $\mathcal{S}_{1,0}(T_{1,d}) \neq \emptyset$, $\mathcal{S}_{1,0}(T_{1,d}) \subseteq \{1, \ldots, n_1\}$. Furthermore, the fundamental detection signature (see (25)) is $\mathcal{S}(T_{1,d}) = \{1\}$. These faults are referred to as *local faults*.

2) Distributed fault, non-overlapping signature: As shown in Fig. 3b, a more general situation arises when links and variables in more than one subsystem are affected by the same single fault, ϕ_2, for which it holds $\mathcal{U}_2 = \{2, 3\}$. This means that, if all LFDs detect a fault at discrete-time instants $T_{I,d}$, $I = 2, 3$, then $\mathcal{S}_{I,0}(T_{I,d}) \neq \emptyset$, $\mathcal{S}_{I,0}(T_{I,d}) \subseteq \{1, \ldots, n_I\}$ and $\mathcal{S}(\bar{t}) = \{2, 3\}$, $\forall \bar{t} \geq \max\{T_{I,d}, I = 2, 3\}$. Furthermore, since there are no shared variables, the local detection signatures are such that $\bigcap_{I \in \{1, \ldots, N\}} \sigma_I(\mathcal{S}_{I,0}(t)) = \emptyset$, $\forall t$.

3) Distributed fault, overlapping signature: A different situation is shown in Fig. 3c where links and variables in more than one subsystem are affected by the same single fault ϕ_3, with $\mathcal{U}_3 = \{1, 2, 3\}$ but, now, shared variables are involved. Specifically, this means that if $T_{I,d}$, $I = 1, 2, 3$, are the local fault detection times of all the LFDs, then $\mathcal{S}_{I,0}(T_{I,d}) \neq$

$\emptyset$, $\mathcal{S}_{I,0}(T_{I,d}) \subseteq \{1, \ldots, n_I\}$ and $\mathcal{S}(\bar{t}) = \{1, 2, 3\}$, $\forall \bar{t} \geq \max\{T_{I,d}, I = 1, 2, 3\}$ and there may exist $\tilde{t} \geq \min\{T_{I,d}, I = 1, 2, 3\}$ such that $\bigcap_{I \in \{1, \ldots, N\}} \sigma_I(\mathcal{S}_{I,0}(t)) \neq \emptyset$, $\forall t \geq \tilde{t}$.

In cases 2) and 3) above, without loss of generality, we considered the situation where all LFDs detect a fault at some finite time. The case where not all LFDs are able to detect a fault can be addressed in an analogous way.

In qualitative and quite rough terms, in this paper, we assume that the generic I–th LFD has access only to the knowledge of the local fault set $\mathcal{F}_I$. Furthermore, the I–th LFD is not informed about the fault influence sets of the global faults corresponding to the local fault functions belonging to $\mathcal{F}_I$. As a consequence, the I–th LFD may only be able to detect and isolate the "*local* part" of a fault that influences the subsystem $\mathscr{S}_I$", but it has not enough information to discern whether the isolated local part correspond to a local fault, or it is just caused by a "larger" distributed fault. This ambiguity is overcome by the third layer (see Fig. 2), consisting of the global fault diagnoser $\mathscr{L}$, which is assumed to have information about the global fault set $\mathcal{F}$ and the fault influence sets of all the global fault functions. By exploiting this knowledge and the local fault decisions d_I^{FD} gathered by all the lower level LFDs, the GFD may be able to take a correct global fault decision d^{FD}: a successful global isolation of a fault by the GFD requires that all of the fault "local parts" have been *locally isolated* by the LFDs in its influence set. In other words, while the goal of each LFD is to locally isolate the local fault among the local fault functions belonging to the local fault set, the task of the GFD is to sort out which one of the global faults has occurred, thus taking a global decision about the health of the monolithic system $\mathscr{S}$.

B. Local fault isolation logic

After a fault has been detected at discrete-time instant T_d and the GFD informs every LFD to switch from the detection to the isolation mode, the FDAE adaptive approximator $\hat{g}_I(t)$ of every LFD stops to learn the interconnection function, that is $\hat{\vartheta}_{I,0}(t) = \hat{\vartheta}_{I,0}(T_d)$, $\forall t \geq T_d$, to prevent the interconnection approximator from keeping on learning also the "influce" of the fault function ϕ_I on the interconnection term. At the same time, each LFD enables its bank of $N_{\mathcal{F}_I}$, $I = 1, \ldots, N$, *Fault Isolation Estimators* (FIEs) in order to implement a GOS for the task of fault isolation, such as the one described in [21]. This scheme relies on the generic l–th FIE of the I–th LFD being matched to the corresponding fault function $\phi_{I,l}$, belonging to the local fault set $\mathcal{F}_I$. Each fault function in $\mathcal{F}_I$ is assumed to be of the form

$$\phi_{I,l}(x_I(t), z_I(t), u_I(t)) = [(\vartheta_{I,l,1})^\top H_{I,l,1}(x_I(t), z_I(t), u_I(t)), \ldots, (\vartheta_{I,l,n_I})^\top H_{I,l,n_I}(x_I(t), z_I(t), u_I(t))]^\top, \quad (26)$$

where, for $k \in \{1, \ldots, n_I\}$, $l \in \{1, \ldots, N_{\mathcal{F}_I}\}$, the *known* functions $H_{I,l,k} : \mathbb{R}^{n_I} \times \mathbb{R}^{p_I} \times \mathbb{R}^{m_I} \mapsto \mathbb{R}^{q_{I,l,k}}$ provide the functional structure of the fault and the *unknown* parameter vectors $\vartheta_{I,l,k} \in \Theta_{I,l,k} \subset \mathbb{R}^{q_{I,l,k}}$ provide its "magnitude". For the sake of simplicity and without much loss of generality, the parameter domains $\Theta_{I,l,k}$ are assumed to be origin–centered hyper–spheres with radius $M_{\Theta_{I,l,k}}$.

After the generic l–th FIE estimator is enabled, with $l \in \{1, \ldots, N_{\mathcal{F}_I}\}$, it monitors its subsystem $\mathscr{S}_I$, providing a *local state estimate* $\hat{x}_{I,l}$ of the local state x_I, analogously to the FDAE. The difference between the estimate $\hat{x}_{I,l}$ and the measurements y_I yields the *estimation error* $\epsilon_{I,l} \triangleq y_I - \hat{x}_{I,l}$ which, again, is used as a residual and compared, component by component, with a suitable *detection threshold* $\bar{\epsilon}_{I,l} \in \mathbb{R}_+^{n_I}$.

The condition

$$|\epsilon_{I,l}^{(k)}(t)| \leq \bar{\epsilon}_{I,l}^{(k)}(t) \ \forall k = 1, \ldots, n_I \qquad (27)$$

is associated to the l–th fault hypothesis

$$\mathscr{H}_{I,l} : \text{"The subsystem } \mathscr{S}_I \text{ is affected by the } l\text{–th fault"}, \qquad (28)$$

with $l = 1, \ldots, N_{\mathcal{F}_I}$. Should condition (27) be violated at some discrete-time instant t, the hypothesis $\mathscr{H}_{I,l}$ is falsified and a so–called *local fault isolation signature* $\mathcal{S}_{I,l}$ is generated.

Definition 16: The l–th *local isolation signature* shown by the subsystem $\mathscr{S}_I$, $I \in \{1, \ldots, N\}$, $l \in \{1, \ldots, N_{\mathcal{F}_I}\}$ at discrete-time instant $t > 0$ is the index set

$$\mathcal{S}_{I,l}(t) \triangleq \{k \in \{1, \ldots, n_I\} : \exists t_1, t \geq t_1 > 0$$
$$\text{such that } |\epsilon_{I,l}^{(k)}(t_1)| > \bar{\epsilon}_{I,l}^{(k)}(t_1)\}. \quad (29)$$

As soon as the hypothesis $\mathscr{H}_{I,l}$ is falsified and the corresponding isolation signature $\mathcal{S}_{I,l}(t)$ becomes non-empty, the specific FIE stops its operation and the fault $\phi_{I,l}(t)$ is excluded as a possible cause of the *detection* signature. The first such time instant is the *exclusion time* $T_{e,I,l}$.

Definition 17: The l–th *fault exclusion time* $T_{e,I,l}$ is defined as $T_{e,I,l} \triangleq \min\{t : \mathcal{S}_{I,l}(t) \neq \emptyset\}$.

Ideally, the goal of the isolation logic is to exclude every but one fault, which *may* be said to be *isolated*. To express this in a formal way, the following definition is introduced.

Definition 18: A fault $\phi_{I,p} \in \mathcal{F}_I$ is *locally isolated* at discrete-time instant t iff $\forall l, l \in \{1, \ldots, N_{\mathcal{F}_I}\} \backslash \{p\}, \mathcal{S}_{I,l}(t) \neq \emptyset$ and $\mathcal{S}_{I,p}(t) = \emptyset$. Furthermore $T_{locisol,I,p} \triangleq \max\{T_{e,I,l}, l \in \{1, \ldots, N_{\mathcal{F}_I}\} \backslash \{p\}\}$ is the *local fault isolation time*.

Remark 4: Again we should note that, if a fault has been locally isolated, we can conclude that it actually occurred if we assume a priori that only faults belonging to the set $\mathcal{F}_I$ may occur. Otherwise, it can only be concluded that it cannot be excluded that it occurred.

C. Local fault isolation and Fault Isolation Estimators

Now, the FIEs are described in detail. After the fault $\phi(t)$ has occurred, the state equation of the s_I–th component of the I–th subsystem becomes

$$x_I^{(s_I)}(t+1) = f_I^{(s_I)}(x_I(t), u_I(t)) + g_I^{(s_I)}(x_I(t), z_I(t), u_I(t)) + \beta(t - T_0)\phi^{(s)}(x(t), u(t)).$$

The l–th FIE estimator dynamic equation for the most general case of a distributed fault, with a shared variable, is defined as

$$\hat{x}_{I,l}^{(s_I)}(t+1) = \lambda \{\hat{x}_{I,l}^{(s_I)}(t) - y_I^{(s_I)}(t) + \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[\hat{x}_{J,l}^{(s_J)}(t)$$
$$- \hat{x}_{I,l}^{(s_I)}(t)]\} + \sum_{J \in \mathcal{O}_s} W_s^{(I,J)}[f_J^{(s_J)}(y_J(t), u_J(t)) + \hat{g}_J^{(s_J)}(t)$$
$$+ \hat{\phi}_{J,l}^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{J,l})], \quad (30)$$

where $\hat{\phi}_{J,l}^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{J,l}) \triangleq (\hat{\vartheta}_{J,l,s_J})^\top \times H_{J,l,s_J}(y_J(t), v_J(t), u_J(t))$ is the s_J–th component of a linearly-parameterized function that matches the structure of the l–th fault function $\phi_{J,l}$, and the vector $\hat{\vartheta}_{J,l} \triangleq \text{col}(\hat{\vartheta}_{J,l,k}, k \in \{1, \ldots, n_I\})$ has been introduced.

Analogously to the FDAE case, the parameters vectors are updated according to the learning law:

$$\hat{\vartheta}_{J,l,k}(t+1) = \mathcal{P}_{\hat{\Theta}_{J,l,k}}(\hat{\vartheta}_{J,l,k}(t) + \gamma_{J,l,k}(t)H_{J,l,k}^\top(t)r_{J,l,k}(t+1)),$$
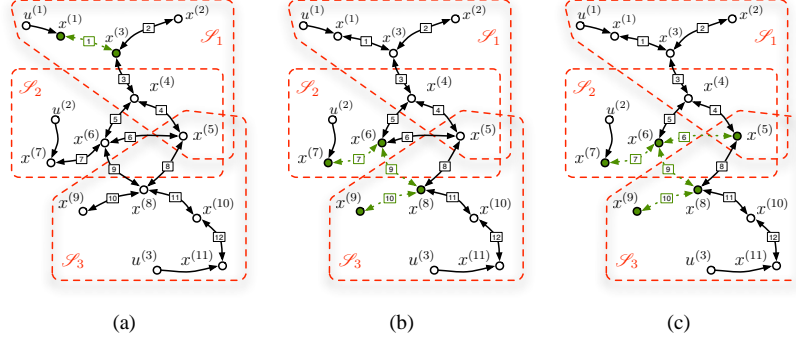
Figure 3. (a) A local fault: for $t \geq T_{1,d}$ only the local detection signature $\mathcal{S}_{1,0}(t)$ of the first LFD is non empty, and the fundamental detection signature is a singleton $\mathcal{S}(t) = \{1\}$, $t \geq T_{1,d}$. (b) A distributed fault with non-overlapping signature: $\forall \bar{t} \geq \max\{T_{I,d}, I = 1,2,3\}$ all the local detection signatures $\mathcal{S}_{I,0}(\bar{t})$ of the LFDs are non empty, and the fundamental detection signature is equal to $\mathcal{S}(\bar{t}) = \{1,2,3\}$ (no shared variables appear in any of the local detection signatures). (c) A distributed fault with overlapping signature: $\forall \bar{t} \geq \max\{T_{I,d}, I = 1,2,3\}$ all the local detection signatures $\mathcal{S}_{I,0}(\bar{t})$ of the LFDs are non empty, and the fundamental detection signature is equal to $\mathcal{S}(\bar{t}) = \{1,2,3\}$ (in this case, shared variables may appear in the local detection signatures).

where $r_{J,l,k}(t+1) = \epsilon_{J,l,k}(t+1) - \lambda\epsilon_{J,l,k}(t)$, and $\mathcal{P}_{\hat{\Theta}_{J,l,k}}$ is again a suitable projection operator

$$\mathcal{P}_{\hat{\Theta}_{J,l,k}}(\hat{\vartheta}_{J,l,k}) \triangleq \begin{cases} \hat{\vartheta}_{J,l,k} & \text{if } |\hat{\vartheta}_{J,l,k}| \leq M_{\hat{\Theta}_{J,l,k}} \\ \frac{M_{\hat{\Theta}_{J,l,k}}}{|\hat{\vartheta}_{J,l,k}|}\hat{\vartheta}_{J,l,k} & \text{if } |\hat{\vartheta}_{J,l,k}| > M_{\hat{\Theta}_{J,l,k}} \end{cases}.$$

The learning rate $\gamma_{J,l,k}(t)$ is computed at each step as $\gamma_{J,l,k}(t) \triangleq \mu_{J,l,k}/(\varepsilon_{J,l,k} + \|H_{J,l,k}^{\top}(t)\|^2)$, with $\varepsilon_{J,l,k} > 0$, $0 < \mu_{J,l,k} < 2$. The corresponding estimation error dynamic equation is

$$\epsilon_{I,l}^{(s_I)}(t+1) = \lambda\{\epsilon_{I,l}^{(s_I)}(t) + \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\epsilon_{J,l}^{(s_J)}(t) - \epsilon_{I,l}^{(s_I)}(t)$$
$$+ \xi_I^{(s_I)}(t) - \xi_J^{(s_J)}(t)]\} + \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\Delta f_J^{(s_J)}(T) + \Delta g_J^{(s_J)}(t)$$
$$+ (1-b^{-(t-T_0)})\phi^{(s)}(t) - \hat{\phi}_{J,l}^{(s_J)}(t)] + \xi_I^{(s_I)}(t+1),$$

which implies

$$\epsilon_{I,l}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\lambda\epsilon_{J,l}^{(s_J)}(t) + \chi_J^{(s_J)}(t)$$
$$+ (1-b^{-(t-T_0)})\phi^{(s)}(t) - \hat{\phi}_{J,l}^{(s_J)}(t)] + \lambda\xi_I^{(s_I)}(t)$$
$$+ \xi_I^{(s_I)}(t+1).$$

Now, considering a matched fault (that is, $\phi^{(s)}(t) = \phi_{J,l}^{(s_J)}$, $\forall J \in \mathcal{O}_s$), the error equation can be written as

$$\epsilon_{I,l}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\lambda\epsilon_{J,l}^{(s_J)}(t) + \chi_J^{(s_J)}(t)$$
$$+ (1-b^{-(t-T_0)})(H_{J,l,s_J}(t)^{\top}\vartheta_{J,l,s_J} + \Delta H_{J,l,s_J}^{\top}\vartheta_{J,l,s_J})$$
$$- H_{J,l,s_J}(t)^{\top}\hat{\vartheta}_{J,l,s_J}] + \lambda\xi_I^{(s_I)}(t) + \xi_I^{(s_I)}(t+1),$$

where $\Delta H_{J,l,s_J}^{\top}(t) \triangleq H_{J,l,s_J}(x_J(t), z_J(t), u_J(t)) - H_{J,l,s_J}(y_J(t), v_J(t), u_J(t))$. By introducing the parameter estimation errors $\tilde{\vartheta}_{J,l,s_J} \triangleq \vartheta_{J,l,s_J} - \hat{\vartheta}_{J,l,s_J}$, the FIE estimation

error equation for a matched fault becomes

$$\epsilon_{I,l}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\lambda\epsilon_{J,l}^{(s_J)}(t) + \chi_J^{(s_J)}(t)$$
$$+ (1-b^{-(t-T_0)})H_{J,l,s_J}(t)^{\top}\tilde{\vartheta}_{J,l,s_J}$$
$$+ (1-b^{-(t-T_0)})\Delta H_{J,l,s_J}(t)^{\top}\vartheta_{J,l,s_J}$$
$$- b^{-(t-T_0)}H_{J,l,s_J}(t)^{\top}\hat{\vartheta}_{J,l,s_J}] + \lambda\xi_I^{(s_I)}(t) + \xi_I^{(s_I)}(t+1),$$

so that its absolute value can be bounded by a threshold that is solution of the following equation

$$\bar{\epsilon}_{I,l}^{(s_I)}(t+1) = \sum_{J\in\mathcal{O}_s} W_s^{(I,J)}[\lambda\bar{\epsilon}_{J,l}^{(s_J)}(t)$$
$$+ \bar{\chi}_J^{(s_J)}(t) + \|H_{J,l,s_J}(t)\|\kappa_{J,l,s_J}(\hat{\vartheta}_{J,l,s_J})$$
$$+ \bar{\Delta}H_{J,l,s_J}(t)\bar{\vartheta}_{J,l,s_J} - \bar{b}^{-(t-T_d)}\|H_{J,l,s_J}(t)\| \|\hat{\vartheta}_{J,l,s_J}\|]$$
$$+ \lambda\bar{\xi}_I^{(s_I)}(t) + \bar{\xi}_I^{(s_I)}(t+1).$$

As in Subsection IV-A, the error and threshold solutions can be conveniently expressed in vector form $\epsilon_{s,l}(t) \triangleq \text{col}(\epsilon_{I,l}^{(s_I)}, I \in \mathcal{O}_s)$, $\bar{\epsilon}_{s,l}(t) \triangleq \text{col}(\bar{\epsilon}_{I,l}^{(s_I)}, I \in \mathcal{O}_s)$, so that it holds

$$\epsilon_{s,l}(t+1) = W_s$$
$$\times [\lambda\epsilon_{s,l}(t) + \chi_s(t) + \text{col}((1-b^{-(t-T_0)})H_{I,l,s_I}(t)^{\top}\tilde{\vartheta}_{I,l,s_I}$$
$$+ (1-b^{-(t-T_0)})\Delta H_{I,l,s_I}(t)^{\top}\vartheta_{I,l,s_I}$$
$$- b^{-(t-T_0)}H_{I,l,s_I}(t)^{\top}\hat{\vartheta}_{I,l,s_I}] + \lambda\xi_s(t) + \xi_s(t+1),$$

$$\epsilon_{s,l}(t) = \sum_{h=T_d}^{t-1} (\lambda W_s)^{t-1-h}W_s$$
$$\times [\chi_s(h) + \text{col}((1-b^{-(h-T_0)})H_{I,l,s_I}(h)^{\top}\tilde{\vartheta}_{I,l,s_I}$$
$$+ (1-b^{-(h-T_0)})\Delta H_{I,l,s_I}(h)^{\top}\vartheta_{I,l,s_I}$$
$$- b^{-(h-T_0)}H_{I,l,s_I}(h)^{\top}\hat{\vartheta}_{I,l,s_I}]$$
$$+ \sum_{h=T_d}^{t-1} [(\lambda W_s)^{t-1-h}(\lambda\xi_s(h) + \xi_s(h+1))] + (\lambda W_s)^{t-T_d}\epsilon_{s,l}(T_d).$$

Componentwise, the estimation error is given by

$$\epsilon_{I,l}^{(s_I)}(t) = w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}$$
$$\times [\chi_{\boldsymbol{s}}(h) + \text{col}((1 - b^{-(h-T_0)})H_{I,l,s_I}(h)^\top \tilde{\vartheta}_{I,l,s_I}$$
$$+ (1 - b^{-(h-T_0)})\Delta H_{I,l,s_I}(h)^\top \vartheta_{I,l,s_I}$$
$$- b^{-(h-T_0)}H_{I,l,s_I}(h)^\top \hat{\vartheta}_{I,l,s_I})]$$
$$+ \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1))]$$
$$+ \lambda \xi_I^{(s_I)}(t-1) + \xi_I^{(s_I)}(t) + \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d}\epsilon_{s,l}(T_d),$$

and, analogously, the threshold solution is given by

$$\bar{\epsilon}_{I,l}^{(s_I)}(t) = w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}$$
$$\times [\bar{\chi}_{\boldsymbol{s}}(t) + \text{col}(\|H_{I,l,s_I}(t)\|\kappa_{I,l,s_I}(\hat{\vartheta}_{I,l,s_I}) + \bar{\Delta}H_{I,l,s_I}(t)\bar{\vartheta}_{I,l,s_I}$$
$$- \bar{b}^{-(t-T_d)}\|H_{I,l,s_I}(t)\| \, \|\hat{\vartheta}_{I,l,s_I}\|)]$$
$$+ \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \bar{\xi}_{\boldsymbol{s}}(t) + \bar{\xi}_{\boldsymbol{s}}(t+1))]$$
$$+ \lambda \bar{\xi}_I^{(s_I)}(t-1) + \bar{\xi}_I^{(s_I)}(t) + \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d}\bar{\epsilon}_{s,l}(T_d).$$

This threshold guarantees by definition that no matched fault is excluded because of uncertainties or the effect of the parameter estimation error $\tilde{\vartheta}_{I,l,s_I}$.

In the case of a non–matched fault (that is, $\phi_I^{(s_I)}(x_I(t), z_I(t), u_I(t)) = \phi_{I,p}^{(s_I)}(x_I(t), z_I(t), u_I(t), \vartheta_{I,p})$ for some $I \in \mathcal{O}_{\boldsymbol{s}}$ and with $p \neq l$), the dynamics of the s_I–component of the estimation error of the l–th FIE of the I–th LFD can be written as

$$\epsilon_{I,l}^{(s_I)}(t+1) = \sum_{J \in \mathcal{O}_{\boldsymbol{s}}} W_{\boldsymbol{s}}^{(I,J)}[\lambda \epsilon_{J,l}^{(s_J)}(t) + \chi_J^{(s_J)}(t)$$
$$+ (1 - b^{-(t-T_0)})\phi_{I,p}^{(s_I)}(x_I(t), z_I(t), u_I(t), \vartheta_{I,p})$$
$$- \hat{\phi}_{J,l}^{(s_J)}(y_J(t), v_J(t), u_J(t), \hat{\vartheta}_{J,l})] + \lambda \xi_I^{(s_I)}(t) + \xi_I^{(s_I)}(t+1).$$

As shown before, a convenient way to study the behavior of the estimation error of the LFDs sharing the variable $x^{(\boldsymbol{s})}$ is to consider the vector $\epsilon_{s,l}$, given by the dynamic equation

$$\epsilon_{s,l}(t+1) = W_{\boldsymbol{s}}[\lambda \epsilon_{s,l}(t) + \chi_{\boldsymbol{s}}(t) + \Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(t)] + \lambda \xi_{\boldsymbol{s}}(t) + \xi_{\boldsymbol{s}}(t+1),$$

where the following *mismatch vector* was introduced

$$\Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(t) \triangleq \text{col}((1 - b^{-(t-T_0)})\phi_{I,p}^{(s_I)}(t), I \in \mathcal{O}_{\boldsymbol{s}}) - \hat{\phi}_{s,l}(t).$$

The solution can then be written as

$$\epsilon_{s,l}(t) = \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h} W_{\boldsymbol{s}}[\chi_{\boldsymbol{s}}(h) + \Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(h)]$$
$$+ \sum_{h=T_d}^{t-1} [(\lambda W_{\boldsymbol{s}})^{t-1-h}(\lambda \xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1))] + (\lambda W_{\boldsymbol{s}})^{t-T_d}\epsilon_{s,l}(T_d),$$

and componentwise is described by

$$\epsilon_{I,l}^{(s_I)}(t) = w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}[\chi_{\boldsymbol{s}}(h) + \Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(h)]$$
$$+ \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1))]$$
$$+ \lambda \xi_I^{(s_I)}(t-1) + \xi_I^{(s_I)}(t) + \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d}\epsilon_{s,l}(T_d).$$

Now, owing to the introduction of the above fault mismatch vector, the following important sufficient condition for *fault isolability* can be proved.

Theorem 2 (Local Fault Isolability): Given a fault $\phi_{I,p} \in \mathcal{F}_I$, if for each $l \in \{1, \ldots, N_{\mathcal{F}_I}\} \setminus \{p\}$ there exists some discrete-time instant $T_l > T_d$ and some $s_I \in \{1, \ldots, n_I\}$ such that

$$\left| w_{\boldsymbol{s},I} \sum_{h=T_d}^{T_l-1} (\lambda W_{\boldsymbol{s}})^{T_l-1-h}\Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(h) \right| >$$
$$w_{\boldsymbol{s},I} \sum_{h=T_d}^{T_l-1} (\lambda W_{\boldsymbol{s}})^{T_l-1-h}[\bar{\chi}_{\boldsymbol{s}}(h) + \text{col}(\|H_{I,l,s_I}(T_l)\|\kappa_{I,l,s_I}(\hat{\vartheta}_{I,l,s_I})$$
$$+ \bar{\Delta}H_{I,l,s_I}(T_l)\bar{\vartheta}_{I,l,s_I} - \bar{b}^{-(T_l-T_d)}\|H_{I,l,s_I}(T_l)\| \, \|\hat{\vartheta}_{I,l,s_I}\|,$$
$$I \in \mathcal{O}_{\boldsymbol{s}})] + 2\left\{ \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{T_l-2} [(\lambda W_{\boldsymbol{s}})^{T_l-2-h}(\lambda \bar{\xi}_{\boldsymbol{s}}(T_l) + \bar{\xi}_{\boldsymbol{s}}(T_l+1))] \right.$$
$$\left. + \lambda \bar{\xi}_I^{(s_I)}(T_l-1) + \bar{\xi}_I^{(s_I)}(T_l) + \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{T_l-1-T_d}\bar{\epsilon}_{s,l}(T_d) \right\},$$

then, the p–th fault is isolated. Furthermore, the local isolation time is upper-bounded by $\max_{l \in \{1, \ldots, N_{\mathcal{F}_I}\} \setminus \{p\}} T_l$.

Proof: By using the triangle inequality, the absolute value of the s_I–th component of the l–th FIE of the I–th LFD estimation error is lower-bounded for $t > T_d$ by

$$|\epsilon_{I,l}^{(s_I)}(t)| \geq \left| w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}\Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(h) \right|$$
$$- \left| w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}\chi_{\boldsymbol{s}}(h) \right|$$
$$- \left| \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \xi_{\boldsymbol{s}}(h) + \xi_{\boldsymbol{s}}(h+1))] \right|$$
$$- |\lambda \xi_{\boldsymbol{s}}^{(I)}(t-1)| - |\xi_{\boldsymbol{s}}^{(I)}(t)| - |\lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d}\epsilon_{s,l}(T_d)|.$$

Using the known bounds on $\gamma_{\boldsymbol{s}}$ and $\xi_{\boldsymbol{s}}$ and the fact that the l–th fault cannot already be excluded at time T_d because of the way its threshold has been defined, we have

$$|\epsilon_{I,l}^{(s_I)}(t)| \geq \left| w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}\Delta_{\boldsymbol{s},l}\phi_{\boldsymbol{s},p}(h) \right|$$
$$- w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h}\bar{\chi}_{\boldsymbol{s}}(h)$$
$$- \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \bar{\xi}_{\boldsymbol{s}}(h) + \bar{\xi}_{\boldsymbol{s}}(h+1))]$$
$$- \lambda \bar{\xi}_{\boldsymbol{s}}^{(I)}(t-1)| - \bar{\xi}_{\boldsymbol{s}}^{(I)}(t) - \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d}\bar{\epsilon}_{s,l}(T_d).$$

In order for the l–th fault to be excluded, the inequality $|\epsilon_{I,l}^{(s_I)}(t)| > \bar{\epsilon}_{I,l}^{(s_I)}(t)$ must be satisfied. This translates to the following further inequality

$$|w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h} \Delta_{\boldsymbol{s},l} \phi_{\boldsymbol{s},p}(h)| \geq$$

$$\bar{\epsilon}_{I,l}(t) + w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-1} (\lambda W_{\boldsymbol{s}})^{t-1-h} \bar{\chi}_{\boldsymbol{s}}(h)$$

$$+ \lambda w_{\boldsymbol{s},I} \sum_{h=T_d}^{t-2} [(\lambda W_{\boldsymbol{s}})^{t-2-h}(\lambda \bar{\xi}_{\boldsymbol{s}}(h) + \bar{\xi}_{\boldsymbol{s}}(h+1))]$$

$$+ \lambda \bar{\xi}_{\boldsymbol{s}}^{(I)}(t-1)| + \bar{\xi}_{\boldsymbol{s}}^{(I)}(t) + \lambda w_{\boldsymbol{s},I}(\lambda W_{\boldsymbol{s}})^{t-1-T_d} \bar{\epsilon}_{\boldsymbol{s},l}(T_d),$$

which is implied by the inequality in the hypothesis of the present theorem. Should the inequality hold for every fault function of $\mathcal{F}_I$ but the p–th, then this fault is locally isolated in the sense of Definition 18. ∎

D. Global fault isolation logic

As discussed earlier, in the proposed DFDI setting a distinction should be drawn on the way local and distributed faults are isolated. If a fault is local, then having the corresponding LFD exclude every but that fault is sufficient for declaring it isolated. However, for distributed faults, the isolation needs that all the LFDs, in the influence set of that fault[11], exclude all other faults. The following formalizes the conditions for a fault, local or distributed, to be globally isolated:

Definition 19: A fault $\phi_{\boldsymbol{l}} \in \mathcal{F}$ is *globally isolated* if for each J-th LFD in the fault influence set $U_{\boldsymbol{l}}$, the corresponding local functions ϕ_{J,l_J} have been isolated, with $J \in \mathcal{U}_{\boldsymbol{l}}$. Furthermore $T_{isol,\boldsymbol{l}} \triangleq \max\{T_{locisol,J,l_J}, J \in \mathcal{U}_{\boldsymbol{l}}\}$ is the *global fault isolation time*.

In practice, the global isolation task is carried out by the GFD, by using the fault influence sets of all the global faults in $\mathcal{F}$, and the LFDs local fault decisions. The GFD isolation logic is detailed in Algorithm 1. In the algorithm, *global isolation* is a boolean variable that is true only when a fault has been successfully globally isolated, while *isolated fault* is the global index of the isolated fault. It is assumed that each LFD sends a fault decision message to the GFD both when it excludes and when it isolates a fault, so that two kinds of message are possible: *excluded* and *isolated*. Clearly, in case of a fault not belonging to the a-priori known fault set, a locally isolated fault may still be excluded at a later discrete-time instant by its LFD.

VI. SIMULATION RESULTS

Re-consider the monolithic system depicted in Fig. 1a (the square labels refer to the pipes number) and decomposed into three overlapping subsystems, according to the decomposition $\mathcal{D} = \{\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_3\}$, with index sets $\mathcal{I}_1 = [1\ 2\ 3\ 4\ 5]^\top$, $\mathcal{I}_2 = [4\ 5\ 6\ 7]^\top$ and $\mathcal{I}_3 = [5\ 8\ 9\ 10\ 11]^\top$. The tank states number 4 and 5 are shared, and the corresponding overlap index sets are $\mathcal{O}_4 = \{1, 2\}$ and $\mathcal{O}_5 = \{2, 3\}$. Three pumps are present, feeding the first, seventh and eleventh tank with the following flows: $u_1 = 1.25 + 0.25 \cdot \sin(0.05 \cdot t)$, $u_2 = 1.9 - 1 \cdot \sin(0.005 \cdot t)$ and $u_3 = 1.3 + 0.6 \cdot \cos(0.03 \cdot t)$. The nominal tank sections are set according to the following vector $A = [1\ 0.5\ 1\ 1\ 2\ 2\ 1\ 1\ 0.5\ 0.5\ 0.5]$ m^2, while the interconnecting pipe cross-sections are nominally equal to $A_p = [0.2\ 0.22\ 0.38\ 0.2\ 0.16\ 0.18\ 0.24\ 0.2\ 0.18\ 0.14\ 0.42\ 0.2]$ m^2.

[11]The fault influence set was introduced in Def. 15.

Algorithm 1 Global fault isolation logic

while $\mathcal{S} = \emptyset$ **do**
    wait for a detection message
end while
notify every LFD to stop learning
notify every LFD to start isolation
global isolation $\leftarrow$ *false*
isolated fault $\leftarrow$ *NULL*
loop
    wait for a local isolation message
    $\boldsymbol{p} \leftarrow$ global fault index corresponding to the fault locally isolated or excluded
    $k \leftarrow$ kind of message
    if $k = excluded$ AND $\boldsymbol{p} = isolated\ fault$ **then**
        global isolation $\leftarrow$ *false*
        isolated fault $\leftarrow$ *NULL*
    else
        if ϕ_{J,p_J} locally isolated for each J such that $J \in \mathcal{U}_{\boldsymbol{p}}$ **then**
            global isolation $\leftarrow$ *true*
            isolated fault $\leftarrow$ $\boldsymbol{p}$
        end if
    end if
end loop

Furthermore, to each tank are connected drain pipes whose nominal cross-section are $A_d = [0.025\ 0.0125\ 0.0225\ 0.0275\ 0.075\ 0.0375\ 0.025\ 0.03\ 0.01\ 0.0125\ 0.015]$ m^2. All the pipes outflow coefficients are unitary. When building the local models f_I of each LFD, the actual cross-sections used are affected by random uncertainties no larger than 5% and 8% of the nominal values, respectively for the tanks and for the pipes. The outflow coefficients are off by no more than 10%. Furthermore the tank levels measurements y_I are affected by measuring uncertainties ξ_I whose components are upper bounded by $\bar{\xi}_1 = [0.05\ 0.05\ 0.05\ 0.05\ 0.05]$ m, $\bar{\xi}_2 = [0.06\ 0.06\ 0.06\ 0.06]$ m, and $\bar{\xi}_3 = [0.04\ 0.04\ 0.04\ 0.04\ 0.04]$ m.

In order to learn the interconnection functions of each subsystem, that in this example account for the flows through pipes crossing a subsystem boundary, each LFD is provided with adaptive approximators $\hat{g}_I$, implemented by RBF neural networks having 3 neurons along the range of each input dimension. The parameter domains Θ_I were chosen to be hyperspheres with radii equal to $[2\ 3\ 2] \cdot T_s$, with $T_s = 0.1$ s being the sampling period. The learning rate auxiliary coefficients for the interconnection adaptive approximators were set to $\mu_{1,0} = 10^{-4}$, $\varepsilon_{1,0} = 10^{-3}$, $\mu_{2,0} = 0.5 \cdot 10^{-4}$, $\varepsilon_{2,0} = 10^{-3}$, $\mu_{3,0} = 0.5 \cdot 10^{-4}$, $\varepsilon_{3,0} = 10^{-3}$, while the filter constants were all set to $\lambda = 0.9$, and the total uncertainties were bounded by $\bar{\chi}_1 = [0.36\ 0.42\ 0.42\ 0.6\ 0.6] \cdot T_s$, $\bar{\chi}_2 = [0.36\ 0.48\ 0.42\ 0.3] \cdot T_s$, $\bar{\chi}_3 = [0.6\ 0.6\ 0.42\ 0.72\ 0.54] \cdot T_s$. The weighting matrices for shared variables were

$$W_4 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}, \quad W_5 = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.2 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}.$$

This can be interpreted, for instance in the case of tank 5, as each of the sharing LFD trusting its own estimate and model three times more than the estimates of every other LFD in the overlap set. Three faults were modelled:

1) **Actuator fault in pump 1, 2 and 3:** partial or full shutdown of all the pumps modelled as $u_f^{(i)} = u^{(i)}(1 - a^{(i)})$, where u_f represents the pumps flow in the faulty

case and $0 \leq a^{(i)} \leq 1$, $i \in \{1, 2, 3\}$.

2) **Leakage in tank 4, 5 and 6:** circular hole of unknown radius $0 \leq \rho^{(i)} \leq A^{(i)}$ in the tank bottom, so that the outflow due to the leak is $q_f^{(i)} = \pi(\rho^{(i)})^2 \sqrt{2gx^{(i)}(t)}$, $i \in \{4, 5, 6\}$.

3) **Breakdown of pipes 3 (tanks 3↔4) and 5 (tanks 4↔6):** partial or complete breakdown of those pipes, so that a relative quota $0 \leq a_p^{(i)} \leq 1$, $i \in \{3, 5\}$ of the water in the pipes is drained out of the tanks instead of flowing between them. This is equivalent to substituting the two pipes with four additional drain pipes, one connected to tank 3, two to tank 4 and one to tank 6.

All these cases represent distributed faults, the fault influence sets being $\mathcal{U}_1 = \mathcal{U}_2 = \{1, 2, 3\}$, $\mathcal{U}_3 = \{1, 2\}$. As can be easily seen, the local fault diagnosers may experience the following local signatures:

- LFD no. 1 can see as local only the breakdown of pump 1, or the leakage in tanks 4 and 5, or the effect on tanks 3 and 4 of the breakdown of pipe 3;
- LFD no. 2 can see as local only the breakdown of pump 2, or the leakage in tanks 4, 5 and 6, or the effect on tanks 4 and 6 of the breakdown of pipe 5;
- LFD no. 3 can see as local only the breakdown of pump 3, or the leakage in tank 5.

The resulting fault sets $\mathcal{F}_I$ are then:

$$\mathcal{F}_1 = \Big\{ [\mathrm{col}(\vartheta_{1,1,1}H_{1,1,1}(t), 0, 0, 0, 0)]^\top,$$
$$[\mathrm{col}(0, 0, 0, \vartheta_{1,2,4}H_{1,2,4}(t), \vartheta_{1,2,5}H_{1,2,5}(t))]^\top,$$
$$[\mathrm{col}(0, 0, \vartheta_{1,3,3}H_{1,3,3}(t), \vartheta_{1,3,4}H_{1,3,4}(t), 0)]^\top \Big\},$$

where $\vartheta_{1,1,1} = a^{(1)}$, $H_{1,1,1}(t) = -\frac{T_s}{A^{(1)}}u_1^{(1)}(t)$, $\vartheta_{1,2,4} = \pi(\rho^{(4)})$, $H_{1,2,4}(t) = -\frac{T_s}{A^{(4)}}\sqrt{2gx_1^{(4)}(t)}$, $\vartheta_{1,2,5} = \pi(\rho^{(5)})$, $H_{1,2,5}(t) = -\frac{T_s}{A^{(5)}}\sqrt{2gx_1^{(5)}(t)}$, $\vartheta_{1,3,3} = a_p^{(3)}$, $H_{1,3,3}(t) = -\frac{T_s}{A^{(3)}}a_p^{(3)}c_p^{(3)}A_p^{(3)} \cdot (\mathrm{sign}(x_1^{(4)}(t) - x_1^{(3)}(t)) \cdot \sqrt{2g|x_1^{(4)}(t) - x_1^{(3)}(t)|} + \sqrt{2gx_1^{(3)}(t)}$, $\vartheta_{1,3,4} = a_p^{(3)}$, $H_{1,3,4}(t) = -\frac{T_s}{A^{(4)}}a_p^{(3)}c_p^{(3)}A_p^{(3)} \cdot (\mathrm{sign}(x_1^{(3)}(t) - x_1^{(4)}(t)) \cdot \sqrt{2g|x_1^{(3)}(t) - x_1^{(4)}(t)|} + \sqrt{2gx_1^{(4)}(t)}$;

$$\mathcal{F}_2 = \Big\{ [\mathrm{col}(0, 0, 0, \vartheta_{2,1,4}H_{2,1,4}(t))]^\top,$$
$$[\mathrm{col}(\vartheta_{2,2,1}H_{2,2,1}(t), \vartheta_{2,2,2}H_{2,2,2}(t), \vartheta_{2,2,3}H_{2,2,3}(t), 0)]^\top,$$
$$[\mathrm{col}(\vartheta_{2,3,1}H_{2,3,1}(t), 0, \vartheta_{2,3,3}H_{2,3,3}(t), 0)]^\top \Big\},$$

where $\vartheta_{2,1,4} = a^{(2)}$, $H_{2,1,4}(t) = -\frac{T_s}{A^{(7)}}u_2^{(1)}(t)$, $\vartheta_{2,2,1} = \pi(\rho^{(4)})$, $H_{2,2,1}(t) = -\frac{T_s}{A^{(4)}}\sqrt{2gx_2^{(1)}(t)}$, $\vartheta_{2,2,2} = \pi(\rho^{(5)})$, $H_{2,2,2}(t) = -\frac{T_s}{A^{(5)}}\sqrt{2gx_2^{(2)}(t)}$, $\vartheta_{2,2,3} = \pi(\rho^{(6)})$, $H_{2,2,3}(t) = -\frac{T_s}{A^{(6)}}\sqrt{2gx_2^{(3)}(t)}$, $\vartheta_{2,3,1} = a_p^{(5)}$, $H_{2,3,1}(t) = -\frac{T_s}{A^{(4)}}a_p^{(5)}c_p^{(5)}A_p^{(5)} \cdot (\mathrm{sign}(x_2^{(3)}(t) - x_2^{(1)}(t)) \cdot \sqrt{2g|x_2^{(3)}(t) - x_2^{(1)}(t)|} + \sqrt{2gx_2^{(1)}(t)}$, $\vartheta_{2,3,3} = a_p^{(5)}$, $H_{2,3,3}(t) = -\frac{T_s}{A^{(6)}}a_p^{(5)}c_p^{(5)}A_p^{(5)} \cdot (\mathrm{sign}(x_2^{(1)}(t) - x_2^{(3)}(t)) \cdot \sqrt{2g|x_2^{(1)}(t) - x_2^{(3)}(t)|} + \sqrt{2gx_2^{(3)}(t)}$;

$$\mathcal{F}_3 = \Big\{ [\mathrm{col}(0, 0, 0, 0, \vartheta_{3,1,5}H_{3,1,5}(t))]^\top,$$
$$[\mathrm{col}(\vartheta_{3,2,1}H_{3,2,1}(t), 0, 0, 0, 0)]^\top \Big\},$$

where $\vartheta_{3,1,5} = a^{(3)}$, $H_{3,1,5}(t) = -\frac{T_s}{A^{(11)}}u_3^{(1)}(t)$, $\vartheta_{3,2,1} = \pi(\rho^{(5)})$, $H_{3,2,1}(t) = -\frac{T_s}{A^{(5)}}\sqrt{2gx_3^{(1)}(t)}$.

Figs. 4–5 show the results of a simulation where at $T_0 = 750$ s an incipient fault of the first kind begins to affect the three pumps, reducing their efficiency by an amount equal, respectively, to 25%, 35% and 20%, with a time constant $b = 1.05$. For each LFD, the detection and isolation residuals components of the three tanks that are directly fed by the pumps, are plotted: tank 1 corresponds to the first local component of subsystem 1, tank 7 to the fourth of subsystem 2, and tank 11 to the fifth of subsystem 3. The sequence of events leading from fault occurrence to fault detection and finally to fault isolation, is summarized in Table VI. A few seconds after the fault occurrence time, the fault is detected by the FDAE of the second LFD, as shown in Fig. 6(a). This results in the second LFD sending a fault detection message to the GFD, that thus computes a non-empty fundamental detection signature. In response to this event, the GFD forces the remaining two LFDs to stop the detection mode, and start the isolation mode of operating. For this reason even if at later times the detection residuals of LFDs number 1 and 3 are able to cross their relative thresholds, these events do not correspond to a fault detection, as the fault was already detected earlier by LFD no. 2. During the isolation mode, all the LFDs are eventually able to reject the fault hypotheses no. 2 and 3, but never the fault hypothesis no. 1, that is thus locally isolated. As the GFD receives the local fault isolation messages from the LFDs, it constantly checks whether for a given fault all the LFDs in its fault influence set have locally isolated it. In the example presented here, fault no. 1 is locally isolated by the third LFD at time 824 s, thus prompting the GFD to globally isolate fault 1 at that same time.

VII. CONCLUSIONS

In this paper, a problem formulation and a distributed fault diagnosis architecture for large-scale dynamical systems was presented. The proposed scheme relies on overlapping decompositions of the system into sets of interconnected simpler subsystems, in order to overcome the scalability issues of a centralized architecture thanks to a *divide et impera* paradigm. Each subsystem is monitored by a local fault diagnosis unit, which is able to detect the presence of faults for the corresponding subsystem based on its own measurements and information from neighboring subsystems. An adaptive approximation scheme is developed in order to learn the functional uncertainty in the interconnection between neighboring subsystems, before any fault is detected. As overlapping decompositions lead to some state components being shared between two or more subsystems, a specially designed consensus-based estimation scheme was devised in order to allow the distributed diagnosis scheme to reach a common decision about faults affecting such variables. Distributed detectability and isolability results were proved in order to show the potential improvements attainable by this consensus scheme w.r.t. a consensus-less one, and in order to provide a way to check the expected sensitivity of the FDI scheme to faults. To the best of the authors knowledge, this is the first work addressing a distributed fault isolation scheme for nonlinear, uncertain large-scale discrete time systems. Simulation results were provided as well to illustrate the effectiveness of the proposed scheme.

Future research effort will be devoted to address several interesting open issues, namely: i) inclusion of time-delays in the dynamic model of the distributed system and in the communication links between the local FDI modules; ii) state variables not available for measurement; iii) validation on

Table I
TIME SEQUENCE OF FAULT OCCURRENCE, DETECTION AND ISOLATION EVENTS.

Time	Event	Detection/Isolation logic results	Figure
750 s	Fault 1 occurs	/	/
752 s	LFD n.2 local detection	$\mathcal{S}_{2,0} = \{4\}$	6(a)
752 s	Global detection	GFD verifies that $\mathcal{S}_0 = \{2\}$	/
752.5 s	LFD n.2 local isolation	$\mathcal{S}_{2,1} = \emptyset$, $\mathcal{S}_{2,2} = \{4\}$, $\mathcal{S}_{2,3} = \{4\}$	6(c,d)
760.5 s	LFD n.1 local isolation	$\mathcal{S}_{1,1} = \emptyset$, $\mathcal{S}_{1,2} = \{1\}$, $\mathcal{S}_{1,3} = \{1\}$	5(c,d)
825 s	LFD n.3 local isolation	$\mathcal{S}_{3,1} = \emptyset$, $\mathcal{S}_{3,2} = \{5\}$	4(c)
825 s	Global isolation	GFD verifies that $\mathcal{S}_{I,2} \neq \emptyset$, $\mathcal{S}_{I,3} \neq \emptyset$ and $\mathcal{S}_{I,1} = \emptyset$, $\forall I \in \mathcal{U}_1$	/

(a)

(b)

(c)
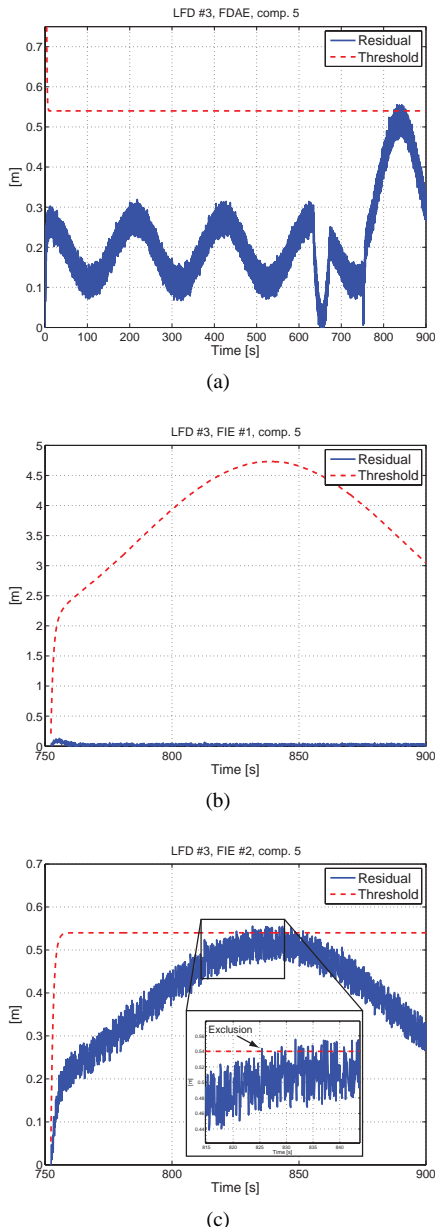
Figure 4. Time–behaviors of simulated signals related to tanks no. 11 when a leakage is introduced at time 750 s. The fault hypothesis no. 2 is locally rejected at time 825 s.

practically-relevant distributed use-cases, both in simulation and in actual experiments. This latter point will require quite significant efforts in order to address implementation issues of the learning algorithms due to the presence of disturbances and variables with different scales. In this connection, it is worth noting that early experiments on a lab-scale experimental setup have shown promising results.

REFERENCES

[1] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, 1998.

[2] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault Tolerant Control*. Berlin: Springer, 2003.

[3] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Berlin: Springer, 2006.

[4] C. Langbort, R. Chandra, and R. D'Andrea, "Distributed control design for systems interconnected over an arbitrary graph," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1502–1519, 2004.

[5] N. Sandell, P. Varaiya, M. Athans, and M. Safonov, "Survey of decentralized control methods for large scale systems," *IEEE Trans. Autom. Control*, vol. 23, no. 2, pp. 108–128, 1978.

[6] P. Baroni, G. Lamperti, P. Pogliano, and M. Zanella, "Diagnosis of large active systems," *Artif. Intell.*, vol. 110, no. 1, pp. 135–189, 1999.

[7] M. Staroswiecki, G. Hoblos, and A. Aitouche, "Sensor network design for fault tolerant estimation," *Int. J. of Adaptive Control and Signal Processing*, vol. 18, no. 1, pp. 55–72, 2004.

[8] R. J. Patton, C. Kambhampati, A. Casavola, P. Zhang, S. Ding, and D. Sauter, "A generic strategy for fault-tolerance in control systems distributed over a network," *Eur. J. Control*, vol. 13, no. 2–3, pp. 280–296, 2007.

[9] S. Klinkhieo and R. J. Patton, "A two-level approach to fault-tolerant control of distributed systems based on the sliding mode," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, Spain, 2009, pp. 1043–1048.

[10] X. G. Yan and C. Edwards, "Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer," *Internat. J. Control*, vol. 81, no. 4, pp. 591–606, 2008.

[11] W. Li, W. Gui, Y. Xie, and S. Ding, "Decentralized fault detection system design for large-scale interconnected systems," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes. Barcelona, Spain*, 2009, pp. 816–821.

[12] X. Zhang, M. M. Polycarpou, and T. Parisini, "Decentralized fault detection in a class of large-scale nonlinear uncertain system," in *Proc. Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, Shanghai, China, 2009, pp. 6988–6993.

[13] N. Meskin, K. Khorasani, and C. A. Rabbath, "Fault consensus in a network of unmanned vehicles," in *Preprints of the 7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, Barcelona, Spain, 2009, pp. 1001–1006.

[14] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "A fault detection scheme for distributed nonlinear uncertain systems," in *Proc. IEEE International Symposium on Intelligent Control*, Munich, Germany, 2006, pp. 2742–2747.

[15] ——, "Distributed fault diagnosis with overlapping decompositions and consensus filters," in *Proc. American Control Conference*, New York, USA, 2007, pp. 693–698.

[16] ——, "Distributed fault diagnosis with overlapping decompositions: an adaptive approximation approach," *IEEE Trans. Autom. Control*, vol. 54, no. 4, pp. 794–799, 2009.

[17] D. Šiljak, *Large-Scale Dynamic Systems: Stability and Structure*. New York: North Holland, 1978.

[18] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault diagnosis of large-scale discrete-time nonlinear systems: New results on the isolation problem," in *Proc. 49th IEEE Conf. on Decision and Control*, Atlanta, USA, 2010, pp. 1619–1626.

[19] M. M. Polycarpou and A. Helmicki, "Automated fault detection and accommodation: a learning systems approach," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 25, no. 11, pp. 1447–1458, 1995.

[20] M. M. Polycarpou and A. Trunov, "Learning approach to nonlinear fault diagnosis: detectability analysis," *IEEE Trans. Autom. Control*, vol. 45, no. 4, pp. 806–812, 2000.

[21] X. Zhang, M. M. Polycarpou, and T. Parisini, "A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems," *IEEE Trans. Autom. Control*, vol. 47, no. 4, pp. 576–593, 2002.

[22] G. Chartrand and O. Oellermann, *Applied and Algorithmic Graph Theory*. Singapore: McGraw-Hill International Editions, 1993.

[23] G. Karypis and V. Kumar, "A fast and high quality multilevel scheme for partitioning irregular graphs," *SIAM J. on Scientific Computing*, vol. 20, no. 1, pp. 359–392, 1999.

[24] X. Cai and Y. Saad, "Overlapping domain decomposition algorithms for general sparse matrices," *Numerical Linear Algebra with Applications*, vol. 3, no. 3, pp. 221–237, 1996.

[25] B. F. Smith, P. Bjorstad, and W. Gropp, *Domain decomposition: parallel multilevel methods for elliptic partial differential equations*. Cambridge, UK: Cambridge University Press, 2004.

[26] J. Lagnese and G. Leugering, *Domain Decomposition Methods in Optimal Control of Partial Differential Equations*. Basel: Birkhäuser, 2004.

[27] H. D. Simon, "Partitioning of unstructured problems for parallel processing," *Computing Systems in Engineering*, vol. 2, no. 2–3, pp. 135–148, 1991.

[28] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[29] D. Johnson, C. Aragon, L. McGeoch, and C. Schevon, "Optimization by simulated annealing: an experimental evaluation. Part I: graph partitioning," *Operations Research*, vol. 37, no. 6, pp. 865–892, 1989.

[30] M. Vidyasagar, "Decomposition techniques for large-scale systems with nonadditive interactions: Stability and stabilizability," *IEEE Trans. Autom. Control*, vol. AC-25, no. 4, pp. 773–779, 1980.

[31] S. Stanković, M. S. Stanković, and D. M. Stipanović, "Consensus based overlapping decentralized estimator," in *Proc. American Control Conference*, New York, USA, 2007, pp. 2744–2749.

[32] M. Ikeda and D. Šiljak, "Overlapping decompositions, expansions, and contractions of dynamic systems," *Large Scale Systems*, vol. 1, no. 1, pp. 29–38, 1980.

[33] M. Hodžič and D. Šiljak, "Decentralized estimation and control with overlapping information sets," *IEEE Trans. Autom. Control*, vol. AC-31, no. 1, 1986.

[34] M. G. Singh, D. Li, Y. Chen, M. Hassan, and Q. Pan, "New approach to failure detection in large-scale systems," *IEE Proceedings D. Control Theory and Applications*, vol. 130, no. 5, pp. 243–249, 1983.

[35] R. D'Andrea and G. E. Dullerud, "Distributed control design for spatially interconnected systems," *IEEE Trans. Autom. Control*, vol. 48, no. 9, pp. 1478–1495, 2003.

[36] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge–based redundancy – A survey and some new results," *Automatica*, vol. 26, no. 3, pp. 459–474, 1990.

[37] R. Patton, P. Frank, and D. Clark, *Fault Diagnosis in Dynamic Systems: Theory and Application*. Upper Saddle River, NJ, USA: Prentice Hall, 1989.

[38] P. Mosterman and G. Biswas, "Diagnosis of continuous valued systems in transient operating regions," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 29, no. 6, pp. 554–565, 1999.

[39] M. M. Polycarpou, "On–line approximators for nonlinear system identification: a unified approach," in *Control and Dynamic Systems: Neural Network Systems Techniques and Applications*, X. Leondes, Ed. New York: Academic, 1998, vol. 7, pp. 191–230.

[40] C. R. Johnson, *Lectures on adaptive parameter estimation*. Upper Saddle River, NJ, USA: Prentice Hall, 1988.

[41] R. Horn and C. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.

[42] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. on Information Process. in Sensor Networks*, Los Angeles, USA, 2005, pp. 63–70.

[43] L. Xiao, S. Boyd, and S. Kim, "Distributed average consensus with least-mean-square deviation," *J. Par. Dist. Comput.*, vol. 67, no. 1, pp. 33–46, 2007.

[44] B. Gharesifard and J. Cortes, "When does a digraph admit a doubly stochastic adjacency matrix?" in *Proc. American Control Conference*, Baltimore, USA, 2010, pp. 2440–2445.

[45] A. Vemuri and M. M. Polycarpou, "On–line approximation methods for robust fault detection," in *Proc. 13th IFAC World Congress*, vol. K, Sidney, AU, 1996, pp. 319–324.

[46] L. Grujic and D. Šiljak, "On stability of discrete composite systems," *IEEE Trans. Autom. Control*, vol. 18, no. 5, pp. 522–524, 1973.

Thomas Parisini received the "Laurea" degree (Cum Laude and printing honours) in Electronic Engineering in 1988 and the Ph.D. degree in Electronic Engineering and Computer Science in 1993 from the University of Genoa. He was with Politecnico di Milano and since 2001 he is Professor and Danieli Endowed Chair of Automation Engineering with University of Trieste. Since 2009, Thomas Parisini is Deputy Rector of University of Trieste and since 2010 he also holds the Chair of Industrial Control at Imperial College London. He authored or co-authored more than 200 research papers in archival journals, book chapters, and international conference proceedings. His research interests include neural-network approximations for optimal control problems, fault diagnosis for nonlinear and distributed systems and nonlinear model predictive control systems. He is a co-recipient of the 2004 Outstanding Paper Award of the IEEE Trans. on Neural Networks and a recipient of the 2007 IEEE Distinguished Member Award. He is involved as Project Leader in several projects funded by the European Union, by the Italian Ministry for Research, and he is currently leading consultancy projects with some major process control companies (ABB, Danieli, Duferco, Electrolux, among others). Thomas Parisini is the Editor-in-Chief of the IEEE Trans. on Control Systems Technology. He was the Chair of the IEEE Control Systems Society Conference Editorial Board, a Distinguished Lecturer of the IEEE Control Systems Society and the Chair of the Technical Committee on Intelligent Control of the IEEE Control Systems Society . He was an elected member of the Board of Governors of the IEEE Control Systems Society and of the European Union Control Association (EUCA) and a member of the board of evaluators of the 7th Framework ICT Research Program of the European Union. Thomas Parisini is currently serving as an Associate Editor of the Int. J. of Control and served as Associate Editor of the IEEE Trans. on Automatic Control, of the IEEE Trans. on Neural Networks, of Automatica, and of the Int. J. of Robust and Nonlinear Control. He was involved in the program and organizing committees of several international conferences. In particular, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and he is General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a Fellow of the IEEE.

Marios M. Polycarpou is a Professor of Electrical and Computer Engineering and the Director of the KIOS Research Center for Intelligent Systems and Networks at the University of Cyprus. He received the B.A. degree in Computer Science and the B.Sc. degree in Electrical Engineering both from Rice University, Houston, TX, USA in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Southern California, Los Angeles, CA, in 1989 and 1992 respectively. In 1992, he joined the University of Cincinnati, Ohio, USA, where he reached the rank of Professor of Electrical and Computer Engineering and Computer Science. In 2001, he was the first faculty to join the newly established Department of Electrical and Computer Engineering at the University of Cyprus, where he served as founding Department Chair from 2001 to 2008. His teaching and research interests are in intelligent systems and control, adaptive and cooperative control systems, computational intelligence, fault diagnosis and distributed agents. Dr. Polycarpou has published more than 200 articles in refereed journals, edited books and refereed conference proceedings, and co-authored the book Adaptive Approximation Based Control, published by Wiley in 2006. He is also the holder of 3 patents. Prof. Polycarpou has served as the Editor-in-Chief of the IEEE Transactions on Neural Networks between 2004-2010. He serves on the Advisory Board of two international journals and is past Associate Editor of the IEEE Transactions on Neural Networks (1998-2003) and of the IEEE Transactions on Automatic Control (1999-2002). He served as the Chair of the Technical Committee on Intelligent Control, IEEE Control Systems Society (2003-05) and as Vice President, Conferences, of the IEEE Computational Intelligence Society (2002-03). Dr. Polycarpou was the recipient of the William H. Middendorf Research Excellence Award at the University of Cincinnati (1997) and was nominated by students for the Professor of the Year award (1996). He is currently an IEEE Distinguished Lecturer in computational intelligence. He participated in more than 50 research projects/grants, funded by several agencies and industry in the United States, by the European Commission and by the Research Promotion Foundation of Cyprus. Dr. Polycarpou is a Fellow of the IEEE and is currently the President-Elect of the IEEE Computational Intelligence Society.
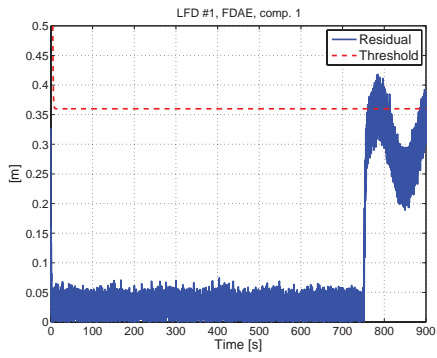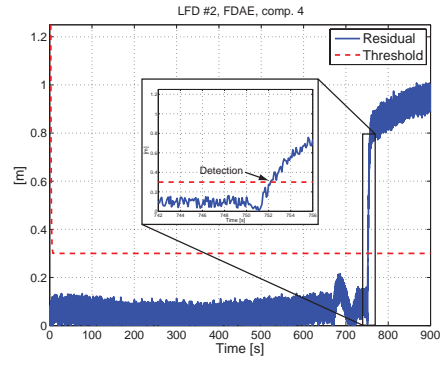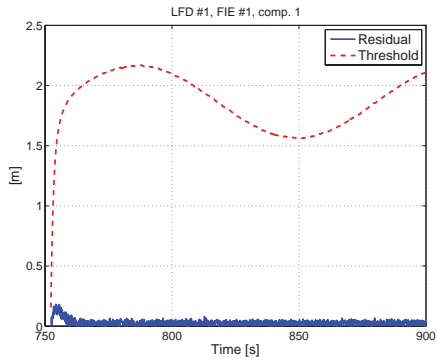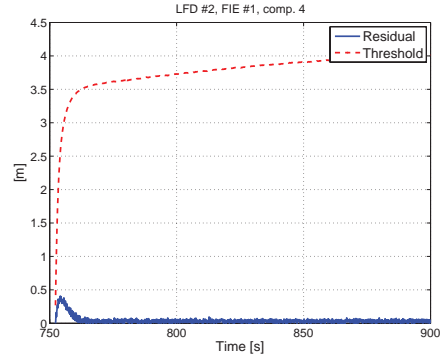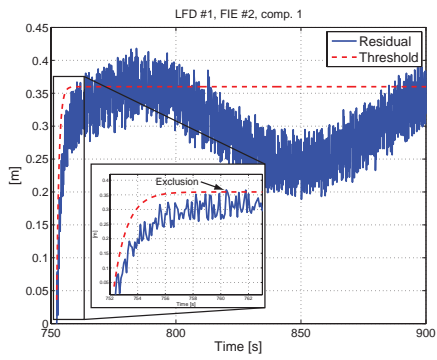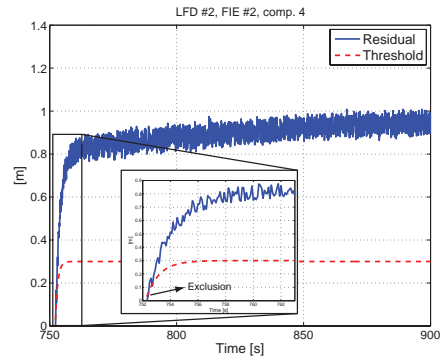
Riccardo M. G. Ferrari received the Laurea (M.Sc.) degree (cum laude) in Electronic Engineering - Control Systems and the Ph.D. degree in Information Science all from University of Trieste, in 2004 and 2009 respectively. He is the recipient of the 2005 Giacomini Award of the Italian Acoustic Society for the best MsC thesis in acoustics. He has authored and co-authored several papers published in international journals and conference proceedings. Since 2008 he is Junior Researcher at the R&D dept. of Danieli Automation S.p.A., Buttrio (UD), Italy. His current research interests include fault diagnosis for nonlinear centralized and distributed dynamic systems, numerical modeling and industrial applications of advanced monitoring and control techniques.
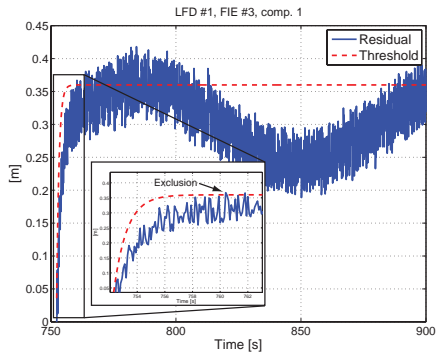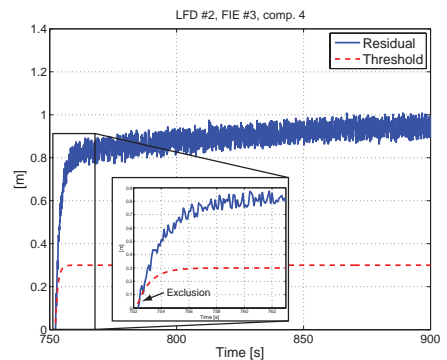
Figure 5. Time–behaviors of simulated signals related to tanks no. 1 when a leakage is introduced at time 750 s. The faults hypotheses no. 2 and 3 are locally rejected shortly after fault detection.

Figure 6. Time–behaviors of simulated signals related to tanks no. 7 when a leakage is introduced at time 750 s. The faults hypotheses no. 2 and 3 are locally rejected shortly after fault detection.