A Distributed Architecture for HVAC Sensor Fault Detection and Isolation

Vasso Reppa, *Member, IEEE,* Panayiotis Papadopoulos, *Student Member, IEEE,*
Marios M. Polycarpou, *Fellow, IEEE,* and Christos G. Panayiotou, *Senior Member, IEEE*

Abstract—This paper presents a design and analysis methodology for detecting and isolating multiple sensor faults in heating, ventilation and air-conditioning (HVAC) systems. The proposed methodology is developed in a distributed framework, considering a multi-zone HVAC system as a set of interconnected, nonlinear subsystems. A dedicated local sensor fault diagnosis (LSFD) agent is designed for each subsystem, while it may exchange information with other LSFD agents. Distributed sensor fault detection is conducted using robust analytical redundancy relations, which are formulated using estimation-based residuals and adaptive thresholds. The distributed sensor fault isolation procedure is carried out by combining the decisions of the LSFD agents and applying a reasoning-based decision logic. The performance of the proposed methodology is analyzed with respect to robustness, sensor fault detectability and isolability. Simulation results are used for illustrating the effectiveness of the proposed methodology applied to an eight-zone HVAC system.

Index Terms—HVAC system, fault detection, fault isolation, sensor faults

I. INTRODUCTION

Recent technological advancements in home automation have contributed to the design of the so-called smart buildings. A smart building can be viewed as a cyber-physical system [1], which consists of the physical-engineered system (the conventional buidling) that is usually large scale and complex, and the cyber core, comprised of communication networks and computational means, designed to monitor, coordinate and control the building environment in order to increase energy efficiency and cost effectiveness, improve comfort, productivity and safety, and increase system robustness and reliability [2], [3]. One of the essential components of a smart building is the Heating, Ventilation and Air Conditioning (HVAC) system, which is responsible for providing a high quality and healthy environment for the building's occupants.

The HVAC system is comprised of a large number of electrical and mechanical components, including the heating and cooling plant (boilers, chillers, dehumidifier), the ventilation system (Variable Air Volume (VAV) terminal units, Air Handling Unit (AHU)), and one or more zones served by the terminal units of the ventilation system. Each subsystem consists of several hardware components, such as sensors (e.g. temperature, humidity), electrical and mechanical actuators

V. Reppa is with the Automatic Control Department, Supélec, Gif-Sur-Yvette, 91400, France, e-mail: vasiliki.reppa@supelec.fr.

P. Papadopoulos, M. M. Polycarpou and C. G. Panayiotou are with the KIOS Research Center for Intelligent Systems and Networks, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, 1678, Cyprus e-mail: ({ppapad01,mpolycar,christosp}@ucy.ac.cy).

(e.g. coils, dampers, valves) and controllers. Over time, it is inevitable that one or more HVAC components will fail, necessitating the utilization of a successful fault detection and isolation (FDI) mechanism [4], [5]. Such mechanism may be one of the enhanced functionalities of a smart building, while, according to [6], it can save 10% to 40% of the HVAC energy consumption.

During the last two decades, various methodologies have been developed for detecting and isolating faults in HVAC systems [7]–[10]. Most of these methodologies have focused on the detection and isolation of faults in actuators and the plant of the HVAC system. However, the detection and isolation of sensor faults is becoming a key challenging problem, since the number of sensors used for monitoring and control of energy consumption and living conditions in large-scale smart buildings is increasing. For example, in the electromechanical part of the HVAC system there may be sensors for measuring supply/return/mixed air temperature, supply/return air flow, differential pressure, return air humidity, etc. Even in a single zone (e.g. room, corridor), there may be a temperature sensor, humidity sensor, CO_2 sensor and an infrared occupancy sensor. Any fault in one or more of these sensors may have significant impact in the smooth operation of the HVAC system, or even jeopardize the safety of the occupants. For example, a fault in the zone temperature sensor (stuck at a high temperature) can cause the continuous operation of the chiller, leading to both discomfort and increased energy consumption; or a fault in the CO_2 sensor can give the wrong signal to the controller for adjusting the air flow of the zone, leading to improper ventilation and unfavorable working conditions.

HVAC sensor faults may also affect the functionality of supervision schemes [11], executing safety critical tasks leading to wrong decisions and disorientation of remedial actions. For example, evacuation plans in case of contaminant release in a building are usually designed in combination with emergency control strategies for the HVAC system. These plans are activated based on measurements of contaminant and occupancy sensors [12]; e.g. aiming at making a zone to be a safe "haven" in case of contamination, the exhaust damper in the zone may be activated for directing the contaminant to another zone, where there are no people, according to the measurements of an occupancy sensor. However, a faulty contaminant sensor may indicate low or zero levels of contamination in the zone, leading to the non-activation of the exhaust damper and characterization of the contaminated zone as safe. Or, the contaminant may be directed by the exhaust damper to a zone, which is indicated as empty, although it is occupied, due to a

faulty occupancy sensor (stuck at zero).

Sensor fault detection and isolation (SFDI) methods for HVAC systems can be classified into data-driven and model-based methods. Data-driven methods are the most commonly used for SFDI in HVAC systems, since they can be developed using a black-box model, without the requirement of understanding the system's model [13]–[18]. However, these methods need a plethora of data collected under both healthy and faulty conditions (data under faulty conditions are necessary for fault isolation), implying increased cost due to the utilization of redundant sensors beyond the sensors required for the proper system operation [19]. On the other hand, model-based methods require additional modeling and calibration effort, since a HVAC model with physical significance has to be developed using a priori knowledge of system process [20]–[23]. Nevertheless, the model-based SFDI methods are designed based on the data acquired by the sensors that are usually installed for feedback control purposes [24]–[27].

The majority of the SFDI methods developed so far are based on a centralized approach, or have focused on the diagnosis of faults in one of the HVAC subsystems, e.g. chiller, AHU, VAV, considering each subsystem separately [16], [17], [24]. HVAC systems are highly complex, nonlinear systems, typically comprised of multiple interconnected subsystems, especially in the case of large-scale buildings, such as hospitals, shopping malls, business centers, airports, universities and many more. Thus, a centralized approach for fault diagnosis may be less suitable compared to a non-centralized approach, since it is characterized by: (i) increased computational complexity of the FDI algorithms, since centralized architectures are tailored to handle (multiple) faults globally, (ii) increased communication requirements due to the transmission of information to a central point, (iii) vulnerability to security threats, because the central cyber core in which the SFDI algorithm resides is a single-point of failure, and (iv) reduced potential of scalability in case of system expansion (e.g. building a new ward in a hospital), due to the utilization of a global physical model or black-box. Moreover, treating the occurrence of faults in a HVAC subsystem separately may be less efficient, since the propagation of faults in a distributed control architecture is neglected. Several researchers have developed decentralized or distributed techniques for diagnosing actuator, process or sensor faults in specific classes of distributed, interconnected nonlinear systems [28]–[34]. However, there are very few distributed techniques for diagnosing multiple sensor faults in HVAC systems [35], which are likely to occur in large-scale buildings.

The objective and main contribution of this work is the design and analysis of a distributed, model-based method for detecting and isolating multiple sensor faults affecting a multi-zone HVAC systems. Based on the nonlinear HVAC model developed in [23], [36], we develop a distributed SFDI methodology exploiting the spatial distribution of the HVAC system; i.e., modeling the HVAC system as a set of $N+1$ interconnected nonlinear systems (N zones and the electromechanical part). For each nonlinear subsystem, we design a dedicated local sensor fault diagnosis (LSFD) agent, which is responsible for detecting and isolating the presence of sensor faults in a distributed manner. To this end, each LSFD agent uses the input and output measurements of its underlying subsystem, as well as the sensor measurements or reference signals of its neighboring subsystems. The sensor fault detection decision logic implemented in the agents relies on checking whether certain analytical redundancy relations (ARRs) are satisfied. The ARRs are formulated using estimation-based residuals and adaptive thresholds, taking into account bounded modeling uncertainties and measurement noise. The distributed isolation of multiple faulty sensors in the HVAC system is carried out using a diagnostic reasoning-based decision logic applied to a sensor fault signature matrix. The performance of the proposed methodology is analyzed with respect to sensor fault detectability and isolability [37], characterizing under certain conditions the class of sensor faults that can be detected and isolated.

The added value of this particular case study is the design of a distributed isolation decision logic and its application to multi-zone HVAC systems that are inherently distributed systems, where the interconnected subsystems are characterized by heterogeneous nonlinear dynamics, as well as the analysis of the different ways that local and propagated sensor faults may affect each subsystem. Moreover, the utilization of adaptive thresholds ensures the robustness of the proposed method against modeling uncertainties and measurement noise, excluding false alarms that are not only annoying to the occupants but also deceptive in emergency situations.

This paper is organized as follows. The HVAC system is described in Section II. The architecture and the design details of the proposed distributed SFDI methodology is presented in Section III. The HVAC sensor fault detectability and isolability are analyzed in Section IV. Simulation results of the application of the proposed SFDI architecture to an eight-zone HVAC system are provided in Section V, followed by concluding remarks in Section VI.

II. HVAC SYSTEM DESCRIPTION

Consider a HVAC system, which consists of N separated zones (e.g. dormitory rooms, classrooms) and the electromechanical part. The basic components of the electromechanical part of the HVAC, shown in Fig. 1 are the cooling coil, the chiller and the chilled water tank, the fan, the supply and return ducts and the variable air volume (VAV) boxes. The cooling coil is connected to the chiller through the chiller water tank, which regulates the water inserted to the cooling coil. The control inputs to the HVAC system are the air flow rate to each of the N zones (controlled through the fan and the VAV boxes) and the chilled water mass flow rate (controlled by a 3-way valve). By controlling these inputs, the objective is to achieve the desired temperature in each building zone (for occupants' comfort) and in the cooling coil (for energy efficiency). The humidity and indoor air quality are not controlled.

The temperature dynamics in each zone, cooling coil and chiller water tank can be modeled based on the fundamental mass and energy conservation equations under the following assumptions [23], [36]: i) the air temperature and velocity have uniform behavior throughout a zone; ii) the transient
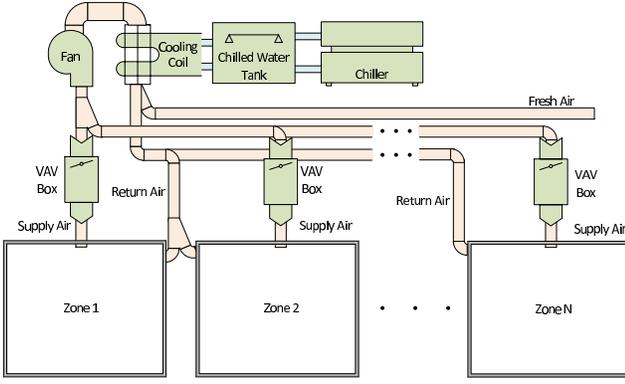
Fig. 1. Schematic of a N-zone HVAC system.

and spatial effects are neglected at the components which exchange air; iii) at the exterior and interior surface of the zones, supply/return ducts, etc., the heat transfer is modeled using constant heat transfer coefficients; iv) the heat transfer at the chilled water tank with the ambient is modeled using a single constant heat transfer coefficient for all surfaces; and v) the axial mixing of water is neglected and the water temperature is constant across the cross section of the tubes.

The temperature dynamic equations of the N-zone HVAC system are described by

$$M_{z_I} C_v \frac{dT_{z_I}(t)}{dt} = \rho_a C_{pa}(T_{ao}(t) - T_{z_I}(t))Q_{a_I}(t)$$
$$+ U_{z_I} A_{z_I}(T_{amb} - T_{z_I}(t)) + \widetilde{T}_{z_I}(t) \tag{1}$$

$$M_{cc} C_v \frac{dT_{ao}(t)}{dt} = \rho_a C_{pa} \left(\frac{1}{N} \sum_{I=1}^{N} T_{z_I}(t) - T_{ao}(t) \right) \sum_{I=1}^{N} Q_{a_I}(t)$$
$$+ U_{cc} A_{cc} \left(T_{amb} - \left(T_{ao}(t) + \frac{1}{N} \sum_{I=1}^{N} T_{z_I}(t) \right) \right)$$
$$+ Q_w \rho_w C_{pw}(T_t(t) - T_{wo})$$
$$+ \rho_a (h_{fg} - C_{pa}) w_z \sum_{I=1}^{N} Q_{a_I}(t)$$
$$- \rho_a (h_{fg} - C_{pa}) w_{ao} \sum_{I=1}^{N} Q_{a_I}(t) \tag{2}$$

$$M_t C_v \frac{dT_t(t)}{dt} = Q_w \rho_w C_{pw}(T_{wo} - T_t(t)) + U_t A_t(T_{amb} - T_t(t))$$
$$+ \frac{15000}{V_t \rho_w C_{pw}} \chi(t), \tag{3}$$

where T_{z_I} (°C) is the temperature of the I-th zone, $I \in \{1, \ldots, N\}$, T_{ao} (°C) is the output air temperature from cooling coil and T_t (°C) is the temperature of the water in the chiller storage tank. The variable Q_{a_I} (m³/ sec) is the volumetric flow rate of air entering into the I-th zone and χ (m³/ sec) is the chilled water mass flow rate. The value $\widetilde{T}_{z_I}(t)$ (°C/sec) represents the rate of internal heat change, due to occupants and appliances from the I-th zone. For the purposes of this paper, it is assumed that the ambient temperature T_{amb} (°C) is constant and known. The remainder constant parameters of the HVAC system are: the heat mass capacitance corresponding to the I-th zone M_{z_I} (kg), specific heat at constant volume C_v (J/kg K), the overall heat transfer

coefficients of the I-th zone, the cooling coil and the chilled water tank U_{z_I}, U_{cc} and U_t (W/m² K), respectively, the density of air and water ρ_a and ρ_w (kg/m³), respectively, the area of the I-th zone, the cooling coil and the chilled water tank A_{z_I}, A_{cc} and A_t (m²), respectively, the specific heat at constant pressure of air and water C_{pa} and C_{pw} (J/kg K), respectively, the latent heat of water h_{fg} (J/kg), the temperature of output water T_{wo} (°C) and the humidity factors w_z, w_{ao} [23].

In each of the N zones, there exist a sensor measuring the zone temperature T_{z_I}, while two sensors are available in the electromechanical part of the HVAC, measuring the temperature of the air exiting the cooling coil T_{ao} and the temperature of the chilled water in the tank T_t. The control inputs to the N-zone HVAC system are the volumetric flow rate of air Q_{a_I} to each zone and the chilled water mass flow rate to the storage tank χ, generated by distributed feedback controllers based on some reference signals. The objective of this work is to design a methodology for detecting and isolating multiple sensor faults that may affect the sensors used for monitoring and control of the N-zone HVAC system.

III. DISTRIBUTED HVAC SENSOR FAULT DETECTION AND ISOLATION

This section provides the design details of the distributed architecture for detection and isolation of sensor faults in the HVAC system described in Section II. The main step for employing the proposed *distributed, model-based* sensor fault diagnosis methodology is to formulate the multi-zone HVAC system given in (1)-(3) as a set of interconnected, nonlinear subsystems, where every local subsystem is described by

$$\dot{x}(t) = Ax(t) + \gamma(x(t), u(t)) + h(x(t), u(t), u_z(t), z(t))$$
$$+ \eta(x(t), u(t), u_z(t), z(t), t), \tag{4}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^\ell$ are the state and input vector of the local subsystem, respectively, while $z \in \mathbb{R}^p$ and $u_z \in \mathbb{R}^{\ell_z}$ are the interconnection state and input vector, containing the states and inputs of the neighboring (interconnected) subsystems. The constant matrix $A \in \mathbb{R}^{n \times n}$ is the linearized part of the state equation and $\gamma : \mathbb{R}^n \times \mathbb{R}^\ell \mapsto \mathbb{R}^n$ represents the known nonlinear dynamics. The term $Ax + \gamma(x, u)$ represents the known local dynamics, while $h : \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^{\ell_z} \times \mathbb{R}^p \mapsto \mathbb{R}^n$ represents the known interconnection dynamics. The last term $\eta : \mathbb{R}^n \times \mathbb{R}^\ell \times \mathbb{R}^{\ell_z} \times \mathbb{R}^p \times \mathbb{R} \mapsto \mathbb{R}^n$ denotes the modeling uncertainty of the local subsystem, representing various sources of uncertainty such as system disturbances, linearization error, uncertainty in the model's parameters, etc. The input vector u is generated by a local feedback controller based a desired reference input.

A. Architecture

The N-zone HVAC system can be regarded as a set of $N + 1$ interconnected, nonlinear subsystems that correspond to the electromechanical part, comprised of the cooling coil and chiller water tank, and the N building zones. Let us define $T^e = [T_1^e, T_2^e]^\top = [T_{ao}, T_t]^\top$, $T_z = [T_{z_1}, \ldots, T_{z_N}]^\top$, $Q_a = [Q_{a_1}, \ldots, Q_{a_N}]^\top$. By writing (2), (3) in the form of (4) with $x \equiv T_e$, $u \equiv \chi$, $z \equiv T_z$, $u_z \equiv Q_a$, the subsystem that

corresponds to the electromechanical part, denoted by Σ^e, can be expressed as:

$$\Sigma^e: \quad \frac{dT^e(t)}{dt} = A^e T^e(t) + \gamma^e(\chi(t)) \\ + h^e(T^e(t), T_z(t), Q_a(t)), \quad (5)$$

where

$$A^e = \begin{bmatrix} -\frac{U_{cc}A_{cc}}{M_{cc}C_v} & \frac{Q_w \rho_w C_{pw}}{M_{cc}C_v} \\ 0 & -\frac{Q_w \rho_w C_{pw} + U_t A_t}{V_t \rho_w C_{pw}} \end{bmatrix} \quad (6)$$

$$\gamma^e(\chi) = \begin{bmatrix} \frac{U_{cc}A_{cc}}{M_{cc}C_v}T_{amb} - \frac{Q_w \rho_w C_{pw}}{M_{cc}C_v}T_{wo} \\ \frac{U_t A_t}{V_t \rho_w C_{pw}}T_{amb} + \frac{Q_w \rho_w C_{pw}}{V_t \rho_w C_{pw}}T_{wo} \end{bmatrix} \\ + \begin{bmatrix} 0 \\ \frac{15000}{V_t \rho_w C_{pw}} \end{bmatrix} \chi, \quad (7)$$

$$h^e(T_1^e, T_z, Q_a) = \begin{bmatrix} h_1^e(T_1^e, T_z, Q_a) \\ 0 \end{bmatrix} \quad (8)$$

$$h_1^e(T_1^e, T_z, Q_a) = \left(\frac{\rho_a C_{pa}}{M_{cc}C_v}\sum_{I=1}^{N}Q_{a_I} - \frac{U_{cc}A_{cc}}{M_{cc}C_v} \right) \frac{1}{N}\sum_{I=1}^{N}T_{z_I} \\ + \frac{\rho_a}{M_{cc}C_v}\left((h_{fg} - C_{pa})(w_z - w_{ao}) \right. \\ \left. - C_{pa}T_1^e \right)\sum_{I=1}^{N}Q_{a_I}. \quad (9)$$

It is noted that the first two terms of (5) represent the local dynamics of Σ^e, while h^e characterizes the interconnection dynamics between Σ^e and $\{\Sigma^{(1)}, \ldots, \Sigma^{(N)}\}$, where $\Sigma^{(I)}$ corresponds to the temperature dynamics of the I-th zone for all $I \in \{1, \ldots, N\}$. By writing (1) in the form of (4) with $x \equiv T_{z_I}$, $u \equiv Q_{a_I}$, $z \equiv T_1^e$, $u_z = 0$, the subsystem of the I-th zone can be expressed as:

$$\Sigma^{(I)}: \quad \frac{dT_{z_I}(t)}{dt} = A^{(I)}T_{z_I}(t) + \gamma^{(I)}(T_{z_I}(t), Q_{a_I}(t)) \\ + h^{(I)}(T_1^e(t), Q_{a_I}(t)) + \eta^{(I)}(t), \quad (10)$$

where $A^{(I)} = -\frac{U_{z_I}A_{z_I}}{M_{z_I}C_v}$, $\eta^{(I)} = \frac{1}{M_{z_I}C_v}\widetilde{T}_{z_I}$ and

$$\gamma^{(I)}(T_{z_I}, Q_{a_I}) = -\frac{\rho_a C_{pa}}{M_{z_I}C_v}T_{z_I}Q_{a_I} + \frac{U_{z_I}A_{z_I}}{M_{z_I}C_v}T_{amb}, \quad (11)$$

$$h^{(I)}(T_1^e, Q_{a_I}) = \frac{\rho_a C_{pa}}{M_{z_I}C_v}T_1^e Q_{a_I}. \quad (12)$$

Again, the first two terms $A^{(I)}T_{z_I}$ and $\gamma^{(I)}(T_{z_I}, Q_{a_I})$ correspond to the local dynamics of $\Sigma^{(I)}$, while $h^{(I)}$ represents the interconnection dynamics between $\Sigma^{(I)}$ and Σ^e.

The I-th subsystem $\Sigma^{(I)}$, $I \in \{1, \ldots, N\}$, is monitored and controlled using a temperature sensor, denoted by $\mathcal{S}^{(I)}$, characterized by the output $y^{(I)} \in \mathbb{R}$; i.e.,

$$\mathcal{S}^{(I)}: \quad y^{(I)}(t) = T_{z_I}(t) + d^{(I)}(t) + f^{(I)}(t), \quad (13)$$

where $d^{(I)} \in \mathbb{R}$ denotes the noise corrupting the measurements $y^{(I)}$ of sensor $\mathcal{S}^{(I)}$ and $f^{(I)} \in \mathbb{R}$ represents the possible sensor fault; i.e., the change in the I-th output $y^{(I)}$ due to a single fault in the I-th sensor is described by:

$$f^{(I)}(t) = \beta^{(I)}(t - t_f^{(I)})\phi^{(I)}(t - t_f^{(I)}), \quad (14)$$

where $\beta^{(I)}$ is the time profile and $\phi^{(I)}$ is the (unknown) function of the sensor fault that occurs at the (unknown) time instant $t_f^{(I)}$. The time profile of the fault is modeled as $\beta^{(I)}(t) = 0$ for $t < 0$ and $\beta^{(I)}(t) = 1 - e^{-k^{(I)}t}$ for $t \geq 0$, where $k^{(I)}$ is the (unknown) evolution rate of the fault. In the case of abrupt sensor faults, the time profile of the fault is modeled by letting $k^{(I)} \to \infty$. In practice, there maybe more than one sensor covering a single zone (especially large zones). In this case, the multiple measurements can be combined by averaging or using advanced sensor fusion methods, while the proposed methodology can still be applied.

The nonlinear subsystem Σ^e is monitored and controlled using a sensor set $\mathcal{S}^e$ that includes two temperature sensors $\mathcal{S}^e\{1\}$ and $\mathcal{S}^e\{2\}$, characterized by

$$\mathcal{S}^e\{1\}: \quad y_1^e(t) = T_1^e(t) + d_1^e(t) + f_1^e(t) \quad (15)$$
$$\mathcal{S}^e\{2\}: \quad y_2^e(t) = T_2^e(t) + d_2^e(t) + f_2^e(t), \quad (16)$$

where $y_j^e \in \mathbb{R}$ is the sensor output, $d_j^e \in \mathbb{R}$ denotes the noise corrupting the measurements of sensor $\mathcal{S}^e\{j\}$ and $f_j^e \in \mathbb{R}$ represents the possible sensor fault described by:

$$f_j^e(t) = \beta_j^e(t - t_{f_j}^e)\phi_j^e(t - t_{f_j}^e), \quad j = 1, 2, \quad (17)$$

where β_j^e is the time profile (the time profile β_j^e is modeled as $\beta^{(I)}$) and ϕ_j^e is the (unknown) function of the sensor fault that occurs at the (unknown) time instant $t_{f_j}^e$. Assuming the occurrence of sensor faults described by (13)-(16) allows us to test several time profiles (the time profile is the way that a fault evolves) and fault functions (the forms of the faults) that may vary for every subsystem.

The design of the proposed distributed SFDI technique is realized as follows. Taking into account the $N+1$ subsystems, defined through (5) and (10), the first step is to design a local sensor fault diagnosis (LSFD) agent for each of the interconnected subsystems; i.e. the agent $\mathcal{M}^e$ dedicated to subsystem Σ^e and the agent $\mathcal{M}^{(I)}$ dedicated to subsystem $\Sigma^{(I)}$, $I \in \{1, \ldots, N\}$ [33], [34], [38]. Each LSFD agent has access to the input and output data of the underlying subsystem, while it may exchange information with some agents. The exchanged information is associated with the form of the physical and input interconnections. Particularly, the agent $\mathcal{M}^e$ that monitors the electromechanical part transmits the measurements of $S^e\{1\}$ to each agent $\mathcal{M}^{(I)}$, while it uses a priori known temperature reference signals of $\Sigma^{(I)}$, $I \in \{1, \ldots, N\}$ from the agent $\mathcal{M}^{(I)}$ [39].

The task of $\mathcal{M}^e$ is to detect and isolate sensor faults affecting $\mathcal{S}^e\{1\}$ and $\mathcal{S}^e\{2\}$. Assuming the occurrence of multiple sensor faults, two modules are designed in the agent $\mathcal{M}^e$ such that the j-th module, denoted by $\mathcal{M}_j^e$ is dedicated to the sensor $\mathcal{S}^e\{j\}$, $j = 1, 2$, and is responsible for isolating a sensor fault that affects $\mathcal{S}^e\{j\}$. The task of $\mathcal{M}^{(I)}$ is to isolate sensor faults in $\mathcal{S}^{(I)}$. However, each agent $\mathcal{M}^{(I)}$ uses the sensor information y_1^e transmitted from $\mathcal{M}^e$, which may be faulty, thus affecting the decision of $\mathcal{M}^{(I)}$; i.e., the agent $\mathcal{M}^{(I)}$ may not be able to distinguish between sensor faults in

both $\mathcal{S}^{(I)}$ and $\mathcal{S}^e\{1\}$. Therefore, the decision of the agent $\mathcal{M}^e$ is transmitted to $\mathcal{M}^{(I)}$ upon request, after the time instant that $\mathcal{M}^{(I)}$ detects the presence of sensor faults [34]. The decision logic implemented in $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \dots, N\}$ relies on checking whether analytical redundancy relations (ARRs) are satisfied, while every ARR is formulated using estimator-based residuals and adaptive thresholds. Taking into account (4), the structure of every estimator, designed for each agent/module, has the following general representation:

$$\dot{\hat{x}}(t) = A x(t) + \gamma(y(t), u(t)) + h(y(t), u(t), u_z(t), z'(t)) + L(y(t) - C\hat{x}(t)) \tag{18}$$

where $\hat{x} \in \mathbb{R}^n$ is the estimation of x (with $\hat{x}(0) = 0$) using the measurements $y \in \mathbb{R}^m$, L is the gain matrix chosen such that the matrix $A - LC$ is stable and $z' \in \mathbb{R}^p$ is comprised of a priori known reference signals or measurements of the interconnection variables z. The sensor output is described by $y(t) = Cx(t) + d(t) + f(t)$, where $C \in \mathbb{R}^{m \times n}$ is the output matrix, while d and f are the noise and fault vector respectively, corrupting the sensor measurements. The estimator (18) is a special case of the Lipschitz observer designed in [33] and [34], satisfying the corresponding assumptions, while the stability of the estimator (18) is ensured if the pair (A, C) is observable.

B. Residual Generation

The first stage of decision-making process conducted by the LSFD agents is the generation of residuals. Residuals are features that portray the status of the monitoring subsystem. Any unusual changes in these features may imply the presence of faults. In this work, residuals represent the deviations of the sensor data (observed behavior) from the estimated sensor outputs (expected behavior).

The nonlinear estimation model of the module $\mathcal{M}_1^e$ is selected as in (18) with $y \equiv y_1^e$, $A \equiv A^e$ $\gamma \equiv \gamma^e$, $h \equiv h^e$ and defining $\widehat{T}_1^e \equiv \hat{x}$; i.e.,

$$\dot{\widehat{T}}_1^e(t) = A^e \widehat{T}_1^e(t) + \gamma^e(\chi(t)) + h^e(y_1^e(t), T_r(t), Q_a(t)) + L_1^e \left(y_1^e(t) - C_1^e \widehat{T}_1^e(t) \right), \tag{19}$$

where $\widehat{T}_1^e \in \mathbb{R}^2$ is the estimation of T^e (using the measurements y_1^e), with initial conditions $\widehat{T}_1^e(0) = [0,0]^\top$, $L_1^e \in \mathbb{R}^{2 \times 1}$ is the estimator gain matrix, chosen such that $A_{L_1}^e = A^e - L_1^e C_1^e$ is stable, $C_1^e = [1,0]$ and $T_r(t) = [T_{r_1}(t), \dots, T_{r_N}(t)]^\top$, where $T_r(t)$ includes the a priori known reference signals of subsystem $\Sigma^{(I)}$, $I \in \{1, \dots, N\}$.

The residual generated by the module $\mathcal{M}_1^e$, is denoted by $\varepsilon_{y_1}^e \in \mathbb{R}$ and is defined as

$$\varepsilon_{y_1}^e(t) = y_1^e(t) - C_1^e \widehat{T}_1^e(t). \tag{20}$$

Let us define the state estimation error $\varepsilon_{T_1}^e(t) = T^e(t) - \widehat{T}_1^e(t)$; given (5), (15) and (19), the residual $\varepsilon_{y_1}^e$ under healthy conditions can be re-written as:

$$\varepsilon_{y_1}^e(t) = C_1^e e^{A_{L_1}^e t} \varepsilon_{T_1}^e(0) + d_1^e(t) + \int_0^t C_1^e e^{A_{L_1}^e (t-\tau)} \bigg(h^e(T_1^e(\tau), T_z(\tau), Q_a(\tau)) - h^e(y_1^e(\tau), T_r(\tau), Q_a(\tau)) - L_1^e d_1^e(\tau) \bigg) d\tau, \tag{21}$$

where y_1^e is the sensor measurement defined in (15). According to (20) and (21), the residual $\varepsilon_{y_1}^e$ is affected only by a possible fault in the sensor $\mathcal{S}^e\{1\}$.

The estimator in the module $\mathcal{M}_2^e$ is structured as in (18) with $y \equiv y_2^e$, $A \equiv A_{22}^e$, $\gamma \equiv \gamma_2^e$ (γ_2^e is the second element of γ^e), $h \equiv 0$ and defining $\widehat{T}_2^e \equiv \hat{x}$; i.e.,

$$\dot{\widehat{T}}_2^e(t) = A_{22}^e \widehat{T}_2^e(t) + \gamma_2^e(\chi(t)) + L_2^e \left(y_2^e(t) - \widehat{T}_2^e(t) \right) \tag{22}$$

where $\widehat{T}_2^e \in \mathbb{R}$ is the estimation of T_2^e, with initial conditions $\widehat{T}_2^e(0) = 0$, A_{22}^e is the element $\{2,2\}$ of the matrix A^e given in (6) and $L_2^e \in \mathbb{R}$ is the estimator gain chosen such $A_{L_2}^e = A_{22}^e - L_2^e$ is stable.

The residual generated by the module $\mathcal{M}_2^e$, denoted by $\varepsilon_{y_2}^e \in \mathbb{R}$, is expressed as:

$$\varepsilon_{y_2}^e(t) = y_2^e(t) - \widehat{T}_2^e(t). \tag{23}$$

where y_2^e is the sensor measurement described by (16). Let us define the state estimation error as $\varepsilon_{T_2}^e(t) = T_2^e(t) - \widehat{T}_2^e(t)$; given (5), (16) and (22), the residual $\varepsilon_{y_2}^e$ under healthy conditions is re-written as:

$$\varepsilon_{y_2}^e(t) = e^{A_{L_2}^e t} \varepsilon_{T_2}^e(0) + d_2^e(t) - \int_0^t e^{A_{L_2}^e (t-\tau)} L_2^e d_2^e(\tau) d\tau. \tag{24}$$

According to (23), (24), the residual $\varepsilon_{y_2}^e$ is affected only by a possible fault in the sensor $\mathcal{S}^e\{2\}$.

The nonlinear estimator implemented in the agent $\mathcal{M}^{(I)}$, $I \in \{1, \dots, N\}$ is structured as in (18) with $y \equiv y^{(I)}$, $A \equiv A^{(I)}$, $\gamma \equiv \gamma^{(I)}$, $h \equiv h^{(I)}$ and defining $\widehat{T}_{z_I} \equiv \hat{x}$; i.e.,

$$\dot{\widehat{T}}_{z_I}(t) = A^{(I)} \widehat{T}_{z_I}(t) + \gamma^{(I)}(y^{(I)}(t), Q_{a_I}(t)) + h^{(I)}(y_1^e(t), Q_{a_I}(t)) + L^{(I)} \left(y^{(I)}(t) - \widehat{T}_{z_I}(t) \right), \tag{25}$$

where $\widehat{T}_{z_I} \in \mathbb{R}$ is the estimation of T_{z_I}, $I \in \{1, \dots, N\}$, with initial conditions $\widehat{T}^{(I)}(0) = 0$ and $L^{(I)} \in \mathbb{R}$ is the estimator gain, chosen such that $A_L^{(I)} = A^{(I)} - L^{(I)}$ is stable; i.e. $L^{(I)} > A^{(I)}$.

The residual generated by the agent $\mathcal{M}^{(I)}$, $I \in \{1, \dots, N\}$, is denoted by $\varepsilon_y^{(I)} \in \mathbb{R}$ and is described by

$$\varepsilon_y^{(I)}(t) = y^{(I)}(t) - \widehat{T}_{z_I}(t), \tag{26}$$

Taking into account (10), (13) and (25), the residual $\varepsilon_y^{(I)}$, $I \in$

$\{1, \ldots, N\}$ under healthy conditions can be expressed as:

$$
\begin{aligned}
\varepsilon_y^{(I)}(t) = {}& e^{A_L^{(I)}t}\varepsilon_x^{(I)}(0) + d^{(I)}(t) + \int_0^t e^{A_L^{(I)}(t-\tau)}\Bigg(\eta^{(I)}(\tau) \\
& - L^{(I)}d^{(I)}(\tau) + \gamma^{(I)}(T_{z_I}(\tau), Q_{a_I}(\tau)) \\
& - \gamma^{(I)}(y^{(I)}(\tau), Q_{a_I}(\tau)) + h^{(I)}(T_1^e(\tau), Q_{a_I}(\tau)) \\
& - h^{(I)}(y_1^e(\tau), Q_{a_I}(\tau))\Bigg)d\tau,
\end{aligned} \tag{27}
$$

where $y^{(I)}$ and y_1^e are sensor measurements described by (13) and (15), respectively. Based on (26) and (27), the residual $\varepsilon_y^{(I)}$ is affected by possible faults in either sensor $\mathcal{S}^e\{1\}$ or sensor $\mathcal{S}^{(I)}$.

C. Computation of Adaptive Thresholds

Due to the presence of disturbances and sensor measurement noise, the observed behavior is typically not identical to the expected behavior even during the healthy operation of the sensors in the building zones and electromechanical part. For this reason, the residuals are compared to thresholds that are designed to bound the residuals under healthy conditions, ensuring the robustness of the agents $\mathcal{M}^e$ and $\mathcal{M}^{(I)}$, for all I, with respect to various sources of uncertainties. The adaptive thresholds designed in this work are time-varying functions of measured or computable signals. The adaptive nature of the thresholds can contribute in reducing the conservativeness in the decision making compared to fixed thresholds. The adaptive thresholds are computed taking into account the following assumption.

Assumption 1: The modeling uncertainty of $\Sigma^{(I)}$, $I \in \{1, \ldots, N\}$ and the measurement noise of each sensor $\mathcal{S}^{(I)}$ and $\mathcal{S}^e\{j\}$, $j = 1, 2$ are unknown but uniformly bounded; i.e., $|\eta^{(I)}(t)| \leq \overline{\eta}^{(I)}$, $|d^{(I)}(t)| \leq \overline{d}^{(I)}$ and $|d_j^e(t)| \leq \overline{d}_j^e$, where $\overline{\eta}^{(I)}$, $\overline{d}_j^{(I)}$, $\overline{d}_j^e$ are known constant bounds.

The bound $\overline{\eta}^{(I)}$ is commonly used for distinguishing between disturbances and faults [40], while the noise bounds $\overline{d}^{(I)}$ and $\overline{d}_j^e$ correspond to a practical representation of the available knowledge for the sensor noise that is typically provided in a given operation range by sensor manufacturers. It is noted that in the case that time varying bounds $\overline{\eta}^{(I)}(t)$, $\overline{d}^{(I)}(t)$ and $\overline{d}_j^e(t)$ are available, this information can be incorporated into the following procedure without significant difficulties.

The adaptive threshold implemented in the module $\mathcal{M}_j^e$, denoted by $\overline{\varepsilon}_{y_j}^e(t)$, $j = 1, 2$, is computed such that

$$|\varepsilon_{y_j}^e(t)| \leq \overline{\varepsilon}_{y_j}^e(t), \tag{28}$$

where $\varepsilon_{y_j}^e(t)$ is the residual defined in (20) and (23). Taking into account Assumption 1 and that there exists a known bound $\overline{T}^e$ such that $|T^e(0)| \leq \overline{T}^e$, and positive constants ρ_1^e, ξ_1^e such that $|C_1^e e^{A_{L_1}^e t}| \leq \rho_1^e e^{-\xi_1^e t}$ for all t, the adaptive threshold is obtained taking into account (21) under healthy conditions ($f_1^e(t) = 0$) and Assumption 1; i.e.,

$$
\begin{aligned}
\overline{\varepsilon}_{y_1}^e(t) = {}& \rho_1^e e^{-\xi_1^e t}\overline{T}^e + \overline{d}_1^e \\
& + \int_0^t \rho_1^e e^{-\xi_1^e(t-\tau)}\left(|L_1^e|\overline{d}_1^e + \overline{h}^e(\tau)\right)d\tau,
\end{aligned} \tag{29}
$$

where $\overline{h}^e(t)$ is computed such that $|h^e(T_1^e(t), T_z(t), Q_a(t)) - h^e(y_1^e(t), T_r(t), Q_a(t))| \leq \overline{h}^e(t)$; i.e.,

$$
\begin{aligned}
\overline{h}^e(t) = {}& \left|\frac{\rho_a C_{pa}}{M_{cc}C_v}\sum_{I=1}^N Q_{a_I}(t) - \frac{U_{cc}A_{cc}}{M_{cc}C_v}\right|\frac{1}{N}\sum_{I=1}^N \overline{T}_I \\
& + \frac{\rho_a C_{pa}}{M_{cc}C_v}\overline{d}_1^e\left|\sum_{I=1}^N Q_{a_I}(t)\right|,
\end{aligned} \tag{30}
$$

where $\overline{T}_I$, is a known constant bound such that $|T_{z_I}(t) - T_{r_I}(t)| \leq \overline{T}_I$, for all t.

Taking into account (24), the adaptive threshold $\overline{\varepsilon}_{y_2}^e$, implemented in the module $\mathcal{M}_2^e$, is described by

$$\overline{\varepsilon}_{y_2}^e(t) = \rho_2^e e^{-\xi_2^e t}\overline{T}_2^e + \overline{d}_2^e + \int_0^t \rho_2^e e^{-\xi_2^e(t-\tau)}|L_2^e|\overline{d}_2^e d\tau, \tag{31}$$

where $\overline{T}_2^e$ is a known bound such that $|T_2^e(0)| \leq \overline{T}_2^e$, and ρ_2^e, ξ_2^e are positive constants such that $|e^{A_{L_2}^e t}| \leq \rho_2^e e^{-\xi_2^e t}$ for all t.

The adaptive threshold implemented in the agent $\mathcal{M}^{(I)}$, denoted by $\overline{\varepsilon}_y^{(I)}(t)$, $I \in \{1, \ldots, N\}$, is computed such that

$$|\varepsilon_y^{(I)}(t)| \leq \overline{\varepsilon}_y^{(I)}(t), \tag{32}$$

where $\varepsilon_y^{(I)}(t)$ is the residual under healthy conditions ($f^{(I)} = 0$, $I \in \{1, \ldots, N\}$ and $f_1^e = 0$) defined in (27). Hence, the adaptive threshold $\overline{\varepsilon}_y^{(I)}(t)$ is described by:

$$
\begin{aligned}
\overline{\varepsilon}_y^{(I)}(t) = {}& \rho^{(I)}e^{-\xi^{(I)}t}\overline{T}_{z_I} + \overline{d}^{(I)} + \int_0^t \rho^{(I)}e^{-\xi^{(I)}(t-\tau)}\Bigg(\overline{\eta}^{(I)} \\
& + \left|L^{(I)}\right|\overline{d}^{(I)} + \frac{\rho_a C_{pa}}{M_{z_I}C_v}\left(\overline{d}^{(I)} + \overline{d}_1^e\right)|Q_{a_I}(\tau)|\Bigg)d\tau,
\end{aligned} \tag{33}
$$

where $\overline{T}_{z_I}$ is a known bound such that $|T_{z_I}(0)| \leq \overline{T}_{z_I}$, $\rho^{(I)}$, $\xi^{(I)}$ are positive constants such that $|e^{A_L^{(I)}t}| \leq \rho^{(I)}e^{-\xi^{(I)}t}$ for all t, and

$$\left|\gamma^{(I)}(T_{z_I}, Q_{a_I}) - \gamma^{(I)}(y^{(I)}, Q_{a_I})\right| \leq \frac{\rho_a C_{pa}}{M_{z_I}C_v}|Q_{a_I}|\overline{d}^{(I)}, \tag{34}$$

$$\left|h^{(I)}(T_1^e, Q_{a_I}) - h^{(I)}(y_1^e, Q_{a_I})\right| \leq \frac{\rho_a C_{pa}}{M_{z_I}C_v}|Q_{a_I}|\overline{d}_1^e. \tag{35}$$

It is noted that the adaptive thresholds defined in (29), (31) and (33) can be implemented using straightforward linear

filtering techniques:

$$\overline{\varepsilon}^e_{y_1} = \rho^e_1 e^{-\xi^e_1 t}\overline{T}^e_1 + \overline{d}^e_1 + H^e_1(s)\left[|L^e_1|\,\overline{d}^e_1 + \overline{h}^e(t)\right], \quad (36)$$

$$\overline{\varepsilon}^e_{y_2} = \rho^e_2 e^{-\xi^e_2 t}\overline{T}^e_2 + \overline{d}^e_2 + H^e_2(s)|L^e_2|\,\overline{d}^e_2, \quad (37)$$

$$\overline{\varepsilon}^{(I)}_y = \rho^{(I)} e^{-\xi^{(I)} t}\overline{T}_{z_I} + \overline{d}^{(I)} + H_I(s)\left(\overline{\eta}^{(I)} + \left|L^{(I)}\right|\overline{d}^{(I)}\right)$$
$$+ H_I(s)\left[\frac{\rho_a C_{pa}}{M_{z_I} C_v}\left(\overline{d}^{(I)} + \overline{d}^e_1\right)|Q_{a_I}(t)|\right], \quad (38)$$

where $H^{(I)}(s) = \frac{\rho^{(I)}}{s+\xi^{(I)}}$, $I \in \{1,\dots,N\}$, $H^e_1(s) = \frac{\rho^e_1}{s+\xi^e_1}$, $H^e_2(s) = \frac{\rho^e_2}{s+\xi^e_2}$ are stable, first-order filters. Note that for any signal $z(t)$, the notation $H(s)[z(t)]$ denotes the output of the filter $H(s)$ with $z(t)$ as input, while s is the Laplace operator.

D. Distributed SFDI Decision Logic

This section presents the decision making process realized by the agent $\mathcal{M}^e$ and its modules $\mathcal{M}^e_1$ and $\mathcal{M}^e_2$, and the agent $\mathcal{M}^{(I)}$, $I \in \{1,\dots,N\}$ for detecting and isolating multiple sensor faults in a distributed manner. The decision logic relies on checking the satisfaction of a set of analytical redundancy relations (ARRs) [41]–[43]. In this work, the ARRs are dynamical constraints, formulated using the residuals and adaptive thresholds.

1) Sensor Fault Detection: The decision logic implemented in the modules $\mathcal{M}^e_1$ and $\mathcal{M}^e_2$, which are included in the agent $\mathcal{M}^e$, is based on the following ARRs:

$$\mathcal{E}^e_j: \quad \left|\varepsilon^e_{y_j}(t)\right| - \overline{\varepsilon}^e_{y_j}(t) \le 0, \; j = 1,2 \quad (39)$$

where $\varepsilon^e_{y_1}$, $\varepsilon^e_{y_2}$ and $\overline{\varepsilon}^e_{y_1}$, $\overline{\varepsilon}^e_{y_2}$ are defined in (21), (24) and (29), (31), respectively. Under healthy conditions, the inequality (39) is always true, implying that the ARRs $\mathcal{E}^e_1$ and $\mathcal{E}^e_2$ are always satisfied. The module $\mathcal{M}^e_j$ infers the presence of sensor fault f^e_j, $j = 1,2$, when $\mathcal{E}^e_j$ defined in (39) is violated. The decision of the module $\mathcal{M}^e_j$, $j = 1,2$ can be described by the following boolean function

$$D^e_j(t) = \begin{cases} 0, & \text{if } t < t^e_{D_j} \\ 1, & \text{if } t \ge t^e_{D_j} \end{cases} \quad (40)$$

$$t^e_{D_j} = \min_t\{t : |\varepsilon^e_{y_j}(t)| - \overline{\varepsilon}^e_{y_j}(t) > 0\} \quad (41)$$

where $t^e_{D_j}$ is the time instant of detection. When $D^e_j(t) = 1$, the module $\mathcal{M}^e_j$, $j = 1,2$ detects the sensor fault f^e_j. Note that as long as $D^e_j(t) = 0$ either there is no sensor fault affecting $\mathcal{S}^e\{j\}$ or sensor fault f^e_j has occurred, but has not been detected by the module $\mathcal{M}^e_j$ until the time instant $t^e_{D_j}$. If $D^e_j(t) = 1$, this implies that the sensor fault f^e_j is guaranteed to affect $\mathcal{S}^e\{j\}$.

The sensor fault detection decision logic of the agent $\mathcal{M}^{(I)}$, $I \in \{1,\dots,N\}$ is based on the following ARR

$$\mathcal{E}^{(I)}: \quad \left|\varepsilon^{(I)}_y(t)\right| - \overline{\varepsilon}^{(I)}_y(t) \le 0, \; I \in \{1,\dots,N\}, \quad (42)$$

where $\varepsilon^{(I)}_y$ and $\overline{\varepsilon}^{(I)}_y$ are defined in (27) and (33), respectively. Under healthy conditions the ARR $\mathcal{E}^{(I)}$, $I \in \{1,\dots,N\}$ is always satisfied. If $\mathcal{E}^{(I)}$ is violated, then this implies that a sensor fault has occurred in either $\mathcal{S}^{(I)}$ or $\mathcal{S}^e\{1\}$ or both of

them. The decision of $\mathcal{M}^{(I)}$ on the presence of sensor faults $f^{(I)}$ or f^e_1 is represented by a boolean function, defined as

$$D^{(I,1)}(t) = \begin{cases} 0, & \text{if } t < t^{(I)}_D \\ 1, & \text{if } t \ge t^{(I)}_D \end{cases} \quad (43)$$

$$t^{(I)}_D = \min_t\{t : |\varepsilon^{(I)}_y(t)| - \overline{\varepsilon}^{(I)}_y(t) > 0\} \quad (44)$$

where $t^{(I)}_D$ is the time of detection for agent $\mathcal{M}^{(I)}$. When $D^{(I,1)}(t) = 1$ the agent $\mathcal{M}^{(I)}$, $I \in \{1,\dots,N\}$ infers that either f^e_1 or $f^{(I)}$ or both, have occurred. As long as $D^{(I,1)}(t) = 0$ either there is no sensor fault in both $\mathcal{S}^{(I)}$ and $\mathcal{S}^e\{1\}$ or sensor faults have occurred, but have not been detected by the agent $\mathcal{M}^{(I)}$ until the time instant $t^{(I)}_D$. If $D^{(I,1)}(t) = 1$, then it is ensured that at least one of $\mathcal{S}^{(I)}$ and $\mathcal{S}^e\{1\}$ is faulty.

2) Sensor Fault Isolation: In the context of smart buildings, it is important not only to be able to detect the occurrence of sensor faults but also to be able to isolate the location of the fault as soon as possible. The agent $\mathcal{M}^e$ can isolate multiple sensor faults in the sensor set $\mathcal{S}^e$ by comparing the observed pattern of sensor faults, defined as $D^e(t) = [D^e_1(t), D^e_2(t)]^\top$ to the columns of the sensor fault signature matrix F^e, presented in Table I. The rows of F^e correspond to the ARRs $\mathcal{E}^e_1$ and $\mathcal{E}^e_2$, while the columns correspond the three possible combinations of sensor faults that occur in $\mathcal{S}^e$, i.e. $\mathcal{F}^e_1 = \{f^e_1\}$, $\mathcal{F}^e_2 = \{f^e_2\}$ and $\mathcal{F}^e_3 = \{f^e_1, f^e_2\}$. The j-th theoretical pattern of the matrix F^e is defined as $F^e_j = \left[F^e_{1j}, F^e_{2j}\right]^\top$, $j = 1,2,3$, where $F^e_{qj} = 1$ if at least one sensor fault of the combination $\mathcal{F}^e_j$ is involved in the ARR $\mathcal{E}^e_q$, and $F^e_{qj} = 0$ otherwise. Based on the sensor fault signature matrix presented in Table I, all possible sensor fault combinations are isolable by the agent $\mathcal{M}^e$, since there are three distinct theoretical patterns.

TABLE I
SENSOR FAULT SIGNATURE MATRIX F^e ($\mathcal{F}^e_1 = \{f^e_1\}$, $\mathcal{F}^e_2 = \{f^e_2\}$ AND $\mathcal{F}^e_3 = \{f^e_1, f^e_2\}$).

	$\mathcal{F}^e_1$	$\mathcal{F}^e_2$	$\mathcal{F}^e_3$
$\mathcal{E}^e_1$	1	0	1
$\mathcal{E}^e_2$	0	1	1

Assuming the occurrence of multiple sensor faults, the decision of the agent $\mathcal{M}^{(I)}$ is combined with the decision of the agent $\mathcal{M}^e$. Specifically, when $\mathcal{M}^{(I)}$ detects the presence of sensor faults ($D^{(I,1)}(t) = 1$), it requests from $\mathcal{M}^e$ to transmit its decision D^e_1 on whether the sensor $\mathcal{S}^e\{1\}$ is faulty in order to isolate the sensor faults. The reason for the combinatorial process of the decisions is that the agent $\mathcal{M}^{(I)}$ uses the measurements of sensor $\mathcal{S}^e_1$ for the generation of the residual and adaptive threshold as well as the formulation of the ARR $\mathcal{E}^{(I)}$. Hence, the distributed sensor fault isolation is conducted by comparing the observed pattern of sensor faults, defined as $D^{(I)}(t) = \left[D^{(I,1)}(t), D^e_1(t)\right]^\top$ to the columns of the sensor fault signature matrix $F^{(I)}$, $I \in \{1,\dots,N\}$, presented in Table II. The rows of $F^{(I)}$ correspond to the ARRs $\mathcal{E}^{(I)}$ and $\mathcal{E}^e_1$, while the columns correspond the three possible combinations of sensor fault occurrence, i.e. $\mathcal{F}^{(I)}_1 = \{f^{(I)}\}$,

$\mathcal{F}_2^{(I)} = \{f_1^e\}$ and $\mathcal{F}_3^{(I)} = \{f^{(I)}, f_1^e\}$.

	$\mathcal{F}_1^{(I)}$	$\mathcal{F}_2^{(I)}$	$\mathcal{F}_3^{(I)}$
$\mathcal{E}^{(I)}$	1	*	1
$\mathcal{E}_1^e$	0	1	1

The j-th column of the matrix $F^{(I)}$ corresponds to the j-th theoretical pattern of sensor faults, defined as $F_j^{(I)} = \left[F_{1j}^{(I)}, F_{2j}^{(I)}\right]^\top$, $j = 1, 2, 3$ where: (i) $F_{qj}^{(I)} = 1$, if the sensor fault combination $\mathcal{F}_j^{(I)}$ contains at least one sensor fault that can provoke the violation of (or else, is involved in) the ARR of the q-th row, $q = 1, 2$ (ii) $F_{qj}^{(I)} = 0$, if none of the sensor faults of the combination $\mathcal{F}_j^{(I)}$ is involved in the ARR of the q-th row, $q = 1, 2$ (iii) $F_{qj}^{(I)} = *$, if none of the sensor faults of the combination $\mathcal{F}_j^{(I)}$ may affect the sensor set $\mathcal{S}^{(I)}$, but all of them are involved in the ARR of the q-th row, $q = 1, 2$. Particularly, the semantics of $F_{21}^{(I)} = *$ implies that the sensor fault f_1^e can explain why $\mathcal{E}^{(I)}$ is violated, but $\mathcal{E}^{(I)}$ may be less sensitive to f_1^e than $f^{(I)}$, so it may be satisfied although f_1^e has occurred. This is based on the fact that the effects of the faulty transmitted information y^e on the residual $\varepsilon_y^{(I)}$ and adaptive threshold $\bar{\varepsilon}_y^{(I)}$, used in the formulation of $\mathcal{E}^{(I)}$, depend on the type of interconnection dynamics $h^{(I)}$, defined in (12). The sensitivity of ARRs to sensor faults is analyzed next.

For isolating multiple sensor faults, the agents $\mathcal{M}^e$ and $\mathcal{M}^{(I)}$ check the consistency between the observed patterns $D^e(t)$ and $D^{(I)}(t)$ and the theoretical patterns F^e and $F^{(I)}$, respectively. As long as $D^e(t) = [0, 0]^\top$ and $D^{(I)}(t) = [0, 0]^\top$, no consistency check is realized; otherwise, the result of the consistency test is the determination of the sensor fault diagnosis set, which contains the diagnosed sensor fault combinations. Specifically, the agent $\mathcal{M}^e$ isolates sensor faults in the electromechanical part of HVAC based on the diagnosis set $\mathcal{D}_s^e(t)$, defined as

$$\mathcal{D}_s^e(t) = \left\{\mathcal{F}_{c_i}^e : i \in \mathcal{I}_\mathcal{D}^e(t)\right\}, \tag{45}$$

where $\mathcal{I}_\mathcal{D}^e(t) = \{i : F_i^e = D^e(t), \ i \in \{1, 2, 3\}\}$. The decision of the agent $\mathcal{M}^{(I)}$, $I \in \{1, \dots, N\}$, relies on the diagnosis set $\mathcal{D}_s^{(I)}(t)$, defined as

$$\mathcal{D}_s^{(I)}(t) = \left\{\mathcal{F}_{c_i}^{(I)} : i \in \mathcal{I}_\mathcal{D}^{(I)}(t)\right\}, \tag{46}$$

where $\mathcal{I}_\mathcal{D}^{(I)}(t) = \left\{i : F_i^{(I)} = D^{(I)}(t), \ i \in \{1, 2, 3\}\right\}$. It is noted that $F_{21}^{(I)} = *$ is consistent to either 0 or 1

Remark 3.1: The proposed sensor fault diagnosis methodology has been developed by applying a dedicated scheme with multiple observers, where each observer of an agent/module is driven by a single sensor (like in $\mathcal{M}_1^e$ and $\mathcal{M}_2^e$) or a set of one local sensor and one sensor in the neighboring subsystem (as the observer in $\mathcal{M}^{(I)}$ for all I). The isolation decision logic relies on the fact that the agents/modules are characterized by (i) robustness, i.e the agents are insensitive to modeling uncertainties and measurement noise under healthy conditions, and (ii) structural fault sensitivity, implying that the agents/modules are sensitive to subsets of sensor faults. Particularly, the agent $\mathcal{M}^{(I)}$ is designed to be structurally sensitive to sensor faults $f^{(I)}$ and f_1^e, while the modules $\mathcal{M}_1^e$ and $\mathcal{M}_2^e$ are sensitive to sensor faults f_1^e and f_2^e, respectively. The residuals are generated using an observer driven by a set of sensors, while the adaptive thresholds are designed to bound the residual under healthy conditions. Therefore, when the magnitude of a residual exceeds the corresponding adaptive threshold, this sensor set is isolated as faulty. An alternative decision logic for isolating sensor faults is to infer that there are faults in a specific sensor set, when the magnitudes of all residuals generated by the observer, which is not driven by this sensor set, do not exceed the corresponding thresholds [44]. This decision logic is applied to a generalized scheme of multiple observers or an unknown input observer (UIO) scheme [45], [46]. In the case of multiple sensor faults, the number of observers in a dedicated scheme may be less than the number of observers in a generalized or UIO scheme.

IV. HVAC SENSOR FAULT DETECTABILITY AND ISOLABILITY

The objective of this section is to analyze the performance of the proposed distributed SFDI methodology with respect to the sensor fault detectability and isolability of the agents $\mathcal{M}^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \dots, N\}$. Specifically, certain conditions are derived, under which we characterize the class of sensor faults affecting $\mathcal{S}^e$, $\mathcal{S}^{(I)}$, $I \in \{1, \dots, N\}$ that can be detected and isolated. It is important to note that the class of detectable/isolable sensor faults satisfying these conditions are obtained under *worst-case assumptions*, in the sense that they are valid for any modeling uncertainty and measurement noise satisfying Assumption 1. It is noted that in practice, the modeling uncertainty and measurement noise may not reach the limit (worst-case) of Assumption 1.

A. Electromechanical Sensor Fault Isolability Conditions

The conditions for guaranteeing the isolation of sensor faults f_1^e and f_2^e by the modules $\mathcal{M}_1^e$ and $\mathcal{M}_2^e$, respectively, are stated in the following Lemma.

Lemma 4.1: Consider that the sensor faults f_1^e and f_2^e occur at the time instants $t_{f_1}^e$ and $t_{f_2}^e$, respectively.

(a) The occurrence of a fault in the temperature sensor of the cooling coil $\mathcal{S}^e\{1\}$ is guaranteed to be isolated under worst-case conditions, if there exists a time instant $t^* > t_{f_1}^e$ such that the sensor fault f_1^e satisfies the condition

$$\left| f_1^e(t^*) - \int_{t_{f_1}^e}^{t^*} C_1^e e^{A_{L_1}^e(t^* - \tau)} \left(L_1^e f_1^e(\tau) \right. \right.$$
$$\left. \left. + \begin{bmatrix} \frac{\rho_a C_{pa}}{M_{cc} C_v} \left(\sum_{I=1}^N Q_{a_I}(\tau) \right) f_1^e(\tau) \\ 0 \end{bmatrix} \right) d\tau \right| > 2\bar{\varepsilon}_{y_1}^e(t^*), \tag{47}$$

where $\bar{\varepsilon}_{y_1}^e(t)$ is the adaptive threshold, generated by the module $\mathcal{M}_1^e$.

(b) The occurrence of a fault in the temperature sensor of the chilled water tank $\mathcal{S}^e\{2\}$ is guaranteed to be isolated under worst-case conditions, if there exists a time instant $t^* > t^e_{f_2}$ such that the sensor fault f^e_2 satisfies the condition

$$\left| f^e_2(t^*) - \int_{t^e_{f_2}}^{t^*} e^{A^e_{L_2}(t^*-\tau)} L^e_2 f^e_2(\tau)d\tau \right| > 2\overline{\varepsilon}^e_{y_2}(t^*), \quad (48)$$

where $\overline{\varepsilon}^e_{y_2}(t)$ is the adaptive threshold, generated by the module $\mathcal{M}^e_2$.

Proof: (a) Assume that no fault affects $\mathcal{S}^e\{1\}$, i.e. $f^e_1 = 0$; then using (5) and (19), the state estimation error of the module $\mathcal{M}^e_1$ satisfies

$$\varepsilon^e_{T_1}(t) = e^{A^e_{L_1}t}\varepsilon^e_{T_1}(0) + \int_0^t e^{A^e_{L_1}(t-\tau)}\left(h^e(T^e_1(\tau), T_z(\tau), Q_a(\tau))\right.$$
$$\left. - h^e(T^e_1(\tau) + d^e_1(\tau), T_r(\tau), Q_a(\tau)) - L^e_1 d^e_1(\tau)\right)d\tau. \quad (49)$$

For $t \geq t^e_{f_1}$, the residual $\varepsilon^e_{y_1}$ is described by:

$$\varepsilon^e_{y_1}(t) = C^e_1 e^{A^e_{L_1}(t-t^e_{f_1})}\varepsilon^e_{T_1}(t^e_{f_1}) + d^e_1(t) + f^e_1(t)$$
$$+ \int_{t^e_{f_1}}^t C^e_1 e^{A^e_{L_1}(t-\tau)}\left(-L^e_1 d^e_1(\tau) - L^e_1 f^e_1(\tau)\right.$$
$$+ h^e(T^e_1(\tau), T_z(\tau), Q_a(\tau)) - h^e(T^e_1(\tau)$$
$$\left. + d^e_1(\tau) + f^e_1(\tau), T_r(\tau), Q_a(\tau))\right)d\tau. \quad (50)$$

By adding and subtracting the integral $\int_{t^e_{f_1}}^t C^e_1 e^{A^e_{L_1}(t-\tau)}h^e(T^e_1(\tau) + d^e_1(\tau), T_r(\tau), Q_a(\tau))d\tau$, and using (49), we obtain

$$\varepsilon^e_{y_1}(t) = \varepsilon^e_{y_{1H}}(t) + \varepsilon^e_{y_{1F}}(t), \quad (51)$$

where $\varepsilon^e_{y_{1H}}(t)$ equals to the residual under healthy conditions described by (21) and $\varepsilon^e_{y_{1F}}(t)$ describes the effects of sensor fault f^e_1 on the residual $\varepsilon^e_{y_1}$, defined as:

$$\varepsilon^e_{y_{1F}}(t) = \int_{t^e_{f_1}}^t C^e_1 e^{A^e_{L_1}(t-\tau)}\left(h^e(T^e_1(\tau) + d^e_1(\tau), T_r(\tau), Q_a(\tau))\right.$$
$$\left. - h^e(T^e_1(\tau) + d^e_1(\tau) + f^e_1(\tau), T_r(\tau), Q_a(\tau))\right)d\tau$$
$$+ f^e_1(t) - \int_{t^e_{f_1}}^t C^e_1 e^{A^e_{L_1}(t-\tau)}L^e_1 f^e_1(\tau)d\tau. \quad (52)$$

Taking into account (28) and (51), it yields

$$\left|\varepsilon^e_{y_1}(t)\right| \geq \left|\varepsilon^e_{y_{1F}}(t)\right| - \left|\varepsilon^e_{y_{1H}}(t)\right| \geq \left|\varepsilon^e_{y_{1F}}(t)\right| - \overline{\varepsilon}^e_{y_1}(t). \quad (53)$$

If there exists a time instant t^* such that the effects of sensor fault f^e_1 on the residual $\varepsilon^e_{y_1}$ satisfy the condition $\left|\varepsilon^e_{y_{1F}}(t^*)\right| > 2\overline{\varepsilon}^e_{y_1}(t^*)$, i.e. satisfy (47), then, based on (53), this implies that $\left|\varepsilon^e_{y_1}(t^*)\right| > \overline{\varepsilon}^e_{y_1}(t^*)$ and the violation of the ARR $\mathcal{E}^e_1$. Thus, sensor fault f^e_1 is guaranteed to be isolated by the module $\mathcal{M}^e_1$.

(b) Assume that no fault affects $\mathcal{S}^e\{2\}$, i.e. $f^e_2 = 0$; using (5) and (22), the state estimation error of the module $\mathcal{M}^e_2$ is

$$\varepsilon^e_{T_2}(t) = e^{A^e_{L_2}t}\varepsilon^e_{T_2}(0) - \int_0^t e^{A^e_{L_2}(t-\tau)}L^e_2 d^e_2(\tau)d\tau. \quad (54)$$

For $t \geq t^e_{f_2}$, the residual $\varepsilon^e_{y_2}$ is expressed as:

$$\varepsilon^e_{y_2}(t) = e^{A^e_{L_2}(t-t^e_{f_2})}\varepsilon^e_{T_2}(t^e_{f_2}) + d^e_2(t) + f^e_2(t)$$
$$- \int_{t^e_{f_2}}^t e^{A^e_{L_2}(t-\tau)}L^e_2\left(f^e_2(\tau) + d^e_2(\tau)\right)d\tau. \quad (55)$$

By replacing $\varepsilon^e_{T_2}(t^e_{f_2})$ using (54), we have

$$\varepsilon^e_{y_2}(t) = \varepsilon^e_{y_{2H}}(t) + \varepsilon^e_{y_{2F}}(t), \quad (56)$$

where $\varepsilon^e_{y_{2H}}(t)$ equals to the residual under healthy conditions described by (24) and $\varepsilon^e_{y_{2F}}(t)$ describes the effects of sensor fault f^e_2 on the residual $\varepsilon^e_{y_2}$, defined as:

$$\varepsilon^e_{y_{2F}}(t) = f^e_2(t) - \int_{t^e_{f_2}}^t e^{A^e_{L_2}(t-\tau)}L^e_2 f^e_2(\tau)d\tau \quad (57)$$

Following the same procedure described in (53), if there exists a time instant t^* such that the effects of sensor fault f^e_2 on the residual $\varepsilon^e_{y_2}$ satisfy the condition $\left|\varepsilon^e_{y_{2F}}(t^*)\right| > 2\overline{\varepsilon}^e_{y_2}(t^*)$, i.e., (48) is valid, then it is implied that $\left|\varepsilon^e_{y_2}(t^*)\right| > \overline{\varepsilon}^e_{y_2}(t^*)$ and the ARR $\mathcal{E}^e_2$ is violated. Thus, sensor fault f^e_2 is guaranteed to be isolated by the module $\mathcal{M}^e_2$. ∎

In general, conditions (47) and (48) can be regarded as a figure of merit, characterizing the ability of $\mathcal{M}^e_1$ and $\mathcal{M}^e_2$ to capture the occurrence of sensor fault f^e_1 and f^e_2, respectively. Based on these conditions, we can define the minimum magnitude of sensor fault f^e_1 and f^e_2 that are isolable by the module $\mathcal{M}^e_1$ and $\mathcal{M}^e_2$, respectively. Particularly, if f^e_1 is constant, i.e. $f^e_1 = \theta^e_1$, and at some time instant t^*, the constant sensor fault θ^e_1 satisfies

$$|\theta^e_1| > \frac{2\overline{\varepsilon}^e_{y_1}(t^*)}{|w(t^*)|} \quad (58)$$

where

$$w(t) = 1 - \int_{t^e_{f_1}}^{t^*} C^e_1 e^{A^e_{L_1}(t^*-\tau)}\left(L^{(I)}\right.$$
$$\left. + \begin{bmatrix} \frac{\rho_a C_{pa}}{M_{cc}C_v}\sum_{I=1}^N Q_{a_I}(\tau) \\ 0 \end{bmatrix}\right)d\tau \quad (59)$$

given that $w(t^*) \neq 0$, the module $\mathcal{M}^e_1$ is guaranteed to isolate sensor fault f^e_1. Similarly, if f^e_2 is constant, i.e. $f^e_2 = \theta^e_2$, and at some time instant t^*, the constant sensor fault θ^e_2 satisfies

$$|\theta^e_2| > \frac{2\overline{\varepsilon}^e_{y_2}(t^*)}{\left|1 - \frac{L^e_2}{A^e_{L_2}}\left(1 - e^{A^e_{L_2}(t^*-t^e_{f_2})}\right)\right|}, \quad (60)$$

given that $\left|1 - \frac{L^e_2}{A^e_{L_2}}\left(1 - e^{A^e_{L_2}(t^*-t^e_{f_2})}\right)\right| \neq 0$, the module $\mathcal{M}^e_2$ is guaranteed to isolate sensor fault f^e_2.

Taking into account (58) and (60), we can characterize the minimum isolable magnitude of sensor fault θ^e_j, $j = 1, 2$, with respect to the bound of sensor noise $\overline{d}^e_j$, and the selected design

parameters used for the implementation of the estimator in the module $\mathcal{M}_j^e$ (e.g. L_j^e) and the adaptive thresholds (ρ_j^e, ξ_j^e).

B. Building Zone Sensor Fault Detectability and Isolability Conditions

The conditions for ensuring the detection/isolation of $f^{(I)}$ and f_1^e by the agent $\mathcal{M}^{(I)}$, are stated in the following Lemma.

Lemma 4.2: Consider that the sensor faults f_1^e and $f^{(I)}$ occur at the time instants $t_{f_1}^e$ and $t_f^{(I)}$, respectively.

(a) Let $t_f^{(I)} < t_{f_1}^e$; the occurrence of a fault in the temperature sensor of the I-th zone $\mathcal{S}^{(I)}$ is guaranteed to be isolated under worst-case conditions, if there exists a time instant $t^* \in [t_f^{(I)}, t_{f_1}^e)$ such that the sensor fault $f^{(I)}$ satisfies the condition

$$\left| f^{(I)}(t^*) - \int_{t_f^{(I)}}^{t^*} e^{A_L^{(I)}(t^*-\tau)} \left(L^{(I)} f^{(I)}(\tau) \right. \right.$$
$$\left. \left. - \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f^{(I)}(\tau) \right) d\tau \right| > 2\bar{\varepsilon}_y^{(I)}(t^*). \quad (61)$$

(b) Let $t_{f_1}^e < t_f^{(I)}$; the occurrence of a fault in the temperature sensor of the cooling coil $\mathcal{S}^e\{1\}$ is guaranteed to be detected under worst-case conditions, if there exists a time instant $t^* \in [t_{f_1}^e, t_f^{(I)})$ such that the sensor fault f_1^e satisfies the condition

$$\left| \int_{t_{f_1}^e}^{t^*} e^{A_L^{(I)}(t^*-\tau)} \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f_1^e(\tau) d\tau \right| > 2\bar{\varepsilon}_y^{(I)}(t^*),$$
$$(62)$$

where $\bar{\varepsilon}_y^{(I)}(t)$ is the adaptive threshold, generated by the agent $\mathcal{M}^{(I)}$.

(c) The occurrence of faults in the temperature sensors $\mathcal{S}^{(I)}$ and $\mathcal{S}^e\{1\}$ is guaranteed to be detected under worst-case conditions, if there exists a time instant $t^* \geq \max(t_f^{(I)}, t_{f_1}^e)$ such that the sensor fault $f^{(I)}$ satisfies the condition

$$\left| f^{(I)}(t^*) - \int_{t_{f_1}^e}^{t^*} e^{A_L^{(I)}(t^*-\tau)} \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f_1^e(\tau) d\tau \right.$$
$$- \int_{t_f^{(I)}}^{t^*} e^{A_L^{(I)}(t^*-\tau)} \left(L^{(I)} f^{(I)}(\tau) \right.$$
$$\left. \left. - \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f^{(I)}(\tau) \right) d\tau \right| > 2\bar{\varepsilon}_y^{(I)}(t^*), \quad (63)$$

Proof: (a) Assume that no fault affects $\mathcal{S}^{(I)}$, $I \in \{1,\ldots,N\}$, and $\mathcal{S}^e\{1\}$, i.e. $f^{(I)} = f_1^e = 0$; based on (10) and (25), the state estimation error of the agent $\mathcal{M}^{(I)}$ is

$$\varepsilon_T^{(I)}(t) = e^{A_L^{(I)} t} \varepsilon_T^{(I)}(0) + \int_0^t e^{A_L^{(I)}(t-\tau)} \left(\eta^{(I)}(\tau) - L^{(I)} d^{(I)}(\tau) \right.$$
$$+ \gamma^{(I)}(T_{z_I}(\tau), Q_{a_I}(\tau)) + h^{(I)}(T_1^e(\tau), Q_{a_I}(\tau))$$
$$- \gamma^{(I)}(T_{z_I}(\tau) + d^{(I)}(\tau), Q_{a_I}(\tau))$$
$$\left. - h^{(I)}(T_1^e(\tau) + d_1^e(\tau), Q_{a_I}(\tau)) \right) d\tau. \quad (64)$$

For $t \geq t_f^{(I)}$, the residual $\varepsilon_y^{(I)}$ is expressed as:

$$\varepsilon_y^{(I)}(t) = e^{A_L^{(I)}\left(t-t_f^{(I)}\right)} \varepsilon_T^{(I)}(t_f^{(I)}) + d^{(I)}(t) + f^{(I)}(t)$$
$$+ \int_{t_f^{(I)}}^t e^{A_L^{(I)}(t-\tau)} \left(\eta^{(I)}(\tau) - L^{(I)} d^{(I)}(\tau) \right.$$
$$- L^{(I)} f^{(I)}(\tau) + \gamma^{(I)}(T_{z_I}(\tau), Q_{a_I}(\tau))$$
$$+ h^{(I)}(T_1^e(\tau), Q_{a_I}(\tau))$$
$$- \gamma^{(I)}(T_{z_I}(\tau) + d^{(I)}(\tau) + f^{(I)}(\tau), Q_{a_I}(\tau))$$
$$\left. - h^{(I)}(T_1^e(\tau) + d_1^e(\tau), Q_{a_I}(\tau)) \right) d\tau. \quad (65)$$

After some algebraic manipulation and using (64) leads to

$$\varepsilon_y^{(I)}(t) = \varepsilon_{y_H}^{(I)}(t) + \varepsilon_{y_F}^{(I)}(t), \quad (66)$$

where $\varepsilon_{y_H}^{(I)}(t)$ corresponds to the residual under healthy conditions described by (27) and $\varepsilon_{y_F}^{(I)}(t)$ describes the effects of sensor fault $f^{(I)}$ on the residual $\varepsilon_y^{(I)}$, defined as:

$$\varepsilon_{y_F}^{(I)}(t) = f^{(I)}(t) - \int_{t_f^{(I)}}^t e^{A_L^{(I)}(t-\tau)} \left(L^{(I)} f^{(I)}(\tau) \right.$$
$$\left. - \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f^{(I)}(\tau) \right) d\tau. \quad (67)$$

Taking into account (32) and (66), it yields

$$\left| \varepsilon_y^{(I)}(t) \right| \geq \left| \varepsilon_{y_F}^{(I)}(t) \right| - \left| \varepsilon_{y_H}^{(I)}(t) \right| \geq \left| \varepsilon_{y_F}^{(I)}(t) \right| - \bar{\varepsilon}_y^{(I)}(t). \quad (68)$$

If there exists a time instant t^* such that the effects of sensor fault $f^{(I)}$ on the residual $\varepsilon_y^{(I)}$ satisfy the condition $\left| \varepsilon_{y_F}^{(I)}(t^*) \right| > 2\bar{\varepsilon}_y^{(I)}(t^*)$, implying that (61) is valid, then, using (68), this entails that $\left| \varepsilon_y^{(I)}(t^*) \right| > \bar{\varepsilon}_y^{(I)}(t^*)$, leading to the isolation of sensor fault $f^{(I)}$.

(b) Part (b) of Lemma 4.2 can be proved in a similar way to part (a).

(c) For $t \geq t_{f_1}^e > t_f^{(I)}$, the residual $\varepsilon_y^{(I)}$ is expressed as:

$$\varepsilon_y^{(I)}(t) = e^{A_L^{(I)}\left(t-t_{f_1}^e\right)} \varepsilon_x^{(I)}(t_{f_1}^e) + d^{(I)}(t) + f^{(I)}(t)$$
$$+ \int_{t_{f_1}^e}^t e^{A_L^{(I)}(t-\tau)} \left(\eta^{(I)}(\tau) - L^{(I)} d^{(I)}(\tau) \right.$$
$$- L^{(I)} f^{(I)}(\tau) + \gamma^{(I)}(T_{z_I}(\tau), Q_{a_I}(\tau))$$
$$- \gamma^{(I)}(T_{z_I}(\tau) + d^{(I)}(\tau) + f^{(I)}(\tau), Q_{a_I}(\tau))$$
$$+ h^{(I)}(T_1^e(\tau), Q_{a_I}(\tau))$$
$$\left. - h^{(I)}(T_1^e(\tau) + d_1^e(\tau) + f_1^e(\tau), Q_{a_I}(\tau)) \right) d\tau. \quad (69)$$

The term $\varepsilon_x^{(I)}(t_{f_1}^e)$ is determined through the following equation

$$
\begin{aligned}
\varepsilon_T^{(I)}(t) = {}& e^{A_L^{(I)}\left(t - t_f^{(I)}\right)} \varepsilon_T^{(I)}(t_f^{(I)}) + \int_{t_f^{(I)}}^t e^{A_L^{(I)}(t-\tau)} \Bigg(\eta^{(I)}(\tau) \\
& - L^{(I)} d^{(I)}(\tau) - L^{(I)} f^{(I)}(\tau) + \gamma^{(I)}(T_{z_I}(\tau), Q_{a_I}(\tau)) \\
& - \gamma^{(I)}(T_{z_I}(\tau) + d^{(I)}(\tau) + f^{(I)}(\tau), Q_{a_I}(\tau)) \\
& + h^{(I)}(T_1^e(\tau), Q_{a_I}(\tau)) \\
& - h^{(I)}(T_1^e(\tau) + d_1^e(\tau), Q_{a_I}(\tau)) \Bigg) d\tau.
\end{aligned}
\tag{70}
$$

Using (64) and (70) and after some algebraic manipulation, the effects of sensor faults $f^{(I)}$ and f_1^e for $t \geq t_{f_1}^e > t_f^{(I)}$ are described as:

$$
\begin{aligned}
\varepsilon_{y_F}^{(I)}(t) = {}& f^{(I)}(t) - \int_{t_{f_1}^e}^t e^{A_L^{(I)}(t-\tau)} \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f_1^e(\tau) d\tau \\
& - \int_{t_f^{(I)}}^t e^{A_L^{(I)}(t^*-\tau)} \Bigg(L^{(I)} f^{(I)}(\tau) \\
& - \frac{\rho_a C_{pa}}{M_{z_I} C_v} Q_{a_I}(\tau) f^{(I)}(\tau) \Bigg) d\tau.
\end{aligned}
\tag{71}
$$

If there exists a time instant t^* such that the effects of sensor fault $f^{(I)}$ on the residual $\varepsilon_y^{(I)}$ satisfy the condition $\left|\varepsilon_{y_F}^{(I)}(t^*)\right| > 2\bar{\varepsilon}_y^{(I)}(t^*)$, implying that (63) is valid, then, using (68) and (71), it is implied that $\left|\varepsilon_y^{(I)}(t^*)\right| > \bar{\varepsilon}_y^{(I)}(t^*)$, leading to the detection of sensor faults $f^{(I)}$ and f_1^e. Following the same procedure, it can be proved that (63) is also valid for $t \geq t_f^{(I)} > t_{f_1}^e$. ∎

Using Lemma (4.2), we may characterize the class of sensor faults $f^{(I)}$ and f_1^e that are detectable/isolable by the agent $\mathcal{M}^{(I)}$ with respect to the bounds of modeling uncertainty and measurement noise, as well as the selected design parameters used for the implementation of the estimator of $\mathcal{M}^{(I)}$ (e.g. $L^{(I)}$) and the adaptive thresholds ($\rho^{(I)}, \xi^{(I)}$). During the design, we can simulate various types of faults, i.e. various fault functions and profiles, which may affect a single sensor, and seek the minimum fault magnitude that satisfies the sensor fault detectability/isolability conditions. This analysis can be performed off-line for calibrating the design parameters before the real-time implementation of the proposed agents.

Comparing (61) to (62), we may infer that sensor fault $f^{(I)}$ affects the residual of $\mathcal{M}^{(I)}$ in a different way than sensor fault f_1^e in the sense that the effects of $f^{(I)}$ are function of $f^{(I)}$ and its filtered version that depends on $L^{(I)}$, while the effects of f_1^e are the filtered version of f_1^e that depends on the interconnection function $h^{(I)}$ only (defined in (12). The fact that sensor fault f_1^e may affect $\mathcal{E}^{(I)}$ in a different way than $f^{(I)}$ is exploited in the design of the sensor fault signature matrix $F^{(I)}$, $I = 1, 2$, by differentiating $F_{11}^{(I)}$ from $F_{12}^{(I)}$. Based on (61) to (62) and assuming constant sensor faults, we may determine the minimum magnitude of sensor fault $f^{(I)}$ and f_1^e that are detectable/isolable by the agent $\mathcal{M}^{(I)}$ in a similar way as in (58) and (60).

For the modules $\mathcal{M}_1^e$ and $\mathcal{M}_2^e$ of the agent $\mathcal{M}^e$, the detectability analysis is equivalent to the isolability analysis, since each module is dedicated to monitor the status of a single sensor, leading to the sensor fault signature matrix presented in Table I. Thus, in Lemma 4.1 we characterize the minimum effects of sensor faults $f_1^e(t)$ and $f_2^e(t)$ that will be isolable by the modules $\mathcal{M}_1^e$ and $\mathcal{M}_2^e$, respectively, by provoking the violation of $\mathcal{E}_1^e$ and $\mathcal{E}_2^e$, respectively. In the case of the agent $\mathcal{M}^{(I)}$, we distinguish the case of a single sensor fault occurrence and the occurrence of two sensor faults. In the first case, we characterize the minimum effects of a local sensor fault ($f^{(I)}$) or a propagated sensor fault (f_1^e) that are guaranteed to be isolable (i.e. provoke the violation of the ARR $\mathcal{E}^{(I)}$) by the agent $\mathcal{M}^{(I)}$ in conjunction with the sensor fault signature matrix presented in Table II. In the second case, we characterize the minimum effects of both local and propagated sensor faults that are guaranteed to provoke the violation of the ARR $\mathcal{E}^{(I)}$.

V. SIMULATION RESULTS

The objective of this section is to illustrate the application of the proposed distributed SFDI method applied to the class of HVAC systems described in Section II consisting of eight zones ($N = 8$) [23]. The operation of the HVAC system is simulated based on equations (5)-(12). The dimensions of each zone are $3.5\text{m} \times 1.75\text{m} \times 2\text{m}$. The parameters used for the simulation of Σ^e described by (5)-(9) are: $\frac{U_{cc} A_{cc}}{M_{cc} C_v}$=0.02815, $\frac{Q_w \rho_w C_{pw}}{M_{cc} C_v}$=1.2084 and $\frac{Q_w \rho_w C_{pw} + U_t A_t}{V_t \rho_w C_{pw}}$=0.0007, T_{wo}=5, $\frac{U_{cc} A_{cc}}{M_{cc} C_v}$=0.02815, $\frac{U_t A_t}{V_t \rho_w C_{pw}}$=5.4566×10^{-4}, $\frac{Q_w \rho_w C_{pw}}{V_t \rho_w C_{pw}}$=1.544×10^{-5} and $\frac{15000}{V_t \rho_w C_{pw}} = 0.006$. The function h^e is defined using the parameters $\frac{\rho_a C_{pa}}{M_{cc} C_v}$=3.932, $\frac{U_{cc} A_{cc}}{M_{cc} C_v}$=0.02815 and $\frac{\rho_a}{M_{cc} C_v}((h_{fg} - C_{pa})(w_z - w_{ao})$=0.0005. The parameters used for the simulation of the subsystem $\Sigma^{(I)}$ $I \in \{1, \ldots, 8\}$ given in (10)-(12) are: $A^{(I)} = -0.0006$, $\frac{\rho_a C_{pa}}{M_{z_I} C_v} = 0.1144$, $\frac{U_{z_I} A_{z_I}}{M_{z_I} C_v} = 0.0006$, $T_{amb} = 35$. The modeling uncertainty $\eta^{(I)}(t)$ is simulated as $\eta^{(I)}(t)$=5%$Y^{(I)}\sin(2\pi\nu t)$, $\nu = 10$ and the noise of each sensor is uniformly distributed, bounded by $\bar{d}^{(j)} = 3\% Y^{(j)}$ and $\bar{d}_j^e = 3\% Y_j^e$, $j = 1, 2$, where $Y^{(I)}, Y_j^e$, are the steady state values of $y^{(I)}, y_j^e$, respectively, under healthy conditions; i.e., controlling the temperatures of each building zone and the electromechanical part and assuming no uncertainty, the steady state values are defined when the temperatures converge to the desired reference signals. Here, $Y_1^e = 10$, $Y_2^e = 4$ and $Y^{(I)} = 24$ for all I.

Eight feedback linearization controllers [47] were implemented, where each controller is responsible for keeping the temperature of each zone at $24°C$. A backstepping controller [48] was applied for maintaining the temperature of the output air of the cooling coil at $10°C$. It is noted that every zone temperature controller uses the measurements of the temperature of the cooling coil, while the controller of the electromechanical part uses the a priori known set points of the temperature of the zones, as well as the air flow rate (control input) of every zone. Based on Section III, we design nine agents, one for the electromechanical part and eight for the zones, while the agent of the electromechanical part consists of two modules. The estimators of the agents are

structured as in (19), (22) and (25) with estimator gains: $L^{(I)} = 3$, $I \in \{1, \ldots, 8\}$, $L_1^e = [4.97, 5.16]^\top$ and $L_2^e = 3$. The adaptive thresholds of the agents, defined in (29), (31) and (33), are designed using the following parameters: $\rho_1^e = 1$, $\xi_1^e = 4$, $\rho_2^e = 1$, $\xi_2^e = 3$ $\rho^{(I)} = 1$ and $\xi^{(I)} = 3$.

We have considered two multiple sensor fault scenarios; in the *first scenario*, the sensors of the electromechanical subsystem and zones 3,4,5,6 are affected by faults, while in the *second scenario*, the sensors in all building zones become faulty. In all scenarios, the sensor faults are abrupt with time varying fault functions; i.e., $\phi_1^e(t) = 15\% Y_1^e + 0.5\sin(0.01t)$, $\phi_2^e(t) = 15\% Y_2^e + 0.5\sin(0.01t)$ and $\phi^{(I)}(t) = 15\% Y^{(I)} + 0.5\sin(0.01t)$, $I \in \{1, \ldots, 8\}$. The time instants of occurrence of sensor faults are: $t_{f_1}^e = 2000$ sec, $t_{f_2}^e = 2500$ sec, $t_f^{(1)} = 3000$ sec, $t_f^{(2)} = 3500$ sec, $t_f^{(3)} = 4000$ sec, $t_f^{(4)} = 4500$ sec, $t_f^{(5)} = 5000$ sec, $t_f^{(6)} = 5500$ sec, $t_f^{(7)} = 6000$ sec and $t_f^{(8)} = 6500$ sec.

The results of the application of the distributed SFDI method to the HVAC system are illustrated in Fig. 2-5, with Fig. 2 and 4 presenting the results for the first sensor fault scenario, while Fig. 3 and 5 for the second scenario. Comparing the observed pattern, $D^e(t) = [D_1^e(t), D_2^e(t)]^\top$, where the temporal evolution of $D_1^e(t)$ and $D_2^e(t)$ is shown in Fig. 2a and 2b, respectively, to the columns of fault signature matrix F^e shown in Table I, the agent $\mathcal{M}^e$ isolates the sensor faults initially in the cooling coil and then in the chilled water tank, based on the following diagnosis set: (i) $\mathcal{D}_s^e(t) = \{f_1^e\}$, since $D^e(t) = F_1^e$ for $t \in [2000, 2500)$, and (ii) $\mathcal{D}_s^e(t) = \{f_1^e, f_2^e\}$, since $D^e(t) = F_3^e$ for $t \geq 2500$.

It is noted that the effects of the sensor fault in the cooling coil on the residuals and thresholds of the eight agents that monitor the building zones are low and are not detectable by these agents (see Fig. 2c-2j). The distinct effects of local sensor fault ($f^{(I)}$) and propagated sensor fault (f_1^e), which are analyzed in Section IV-B, can be observed through the simulation results presented in Fig. 2c-2j. Based on Fig. 2c, 2d, 2i and 2j, the agents $\mathcal{M}^{(1)}$, $\mathcal{M}^{(2)}$, $\mathcal{M}^{(7)}$ and $\mathcal{M}^{(8)}$ do not detect the presence of the faulty temperature sensor in the cooling coil, although they use its measurements. These agents do not also detect the occurrence of sensor faults in the building zones 3,4,5 and 6, but this is due to the fact that every agent $\mathcal{M}^{(I)}$ is sensitive to faults $f^{(I)}$ and f_1^e and not to fault $f^{(Q)}$, $Q \neq I$.

Each of the agents $\mathcal{M}^{(3)}$, $\mathcal{M}^{(4)}$, $\mathcal{M}^{(5)}$ and $\mathcal{M}^{(6)}$ detects the presence of sensor faults just after the consecutive occurrence of the sensor fault in each monitoring building zone, as presented in Fig. 2e, 2f, 2g and 2h. Then, using the decision of $\mathcal{M}_1^e$, the agent $\mathcal{M}^{(I)}$ compares the observed pattern $D^{(I)}(t) = [D^{(I,1)}(t), D_1^e(t)]^\top$, $I \in \{3, 4, 5, 6\}$, where the temporal evolution of D_1^e, $D^{(3,1)}$, $D^{(4,1)}$, $D^{(5,1)}$ and $D^{(6,1)}$ are presented in Fig. 2a, 2e, 2f, 2g and 2h to the columns of the sensor fault signature matrix $F^{(I)}$ shown in Table II. Given that $D^{(I)}(t) = [1, 1]^\top$ for all $I \in \{3, 4, 5, 6\}$, the resultant diagnosis set is $\mathcal{D}^{(I)}(t) = \{\{f^{(I)}, f_1^e\}, f_1^e\}$. Based on this diagnosis outcome, the agent $\mathcal{M}^{(I)}$ for all $I \in \{3, 4, 5, 6\}$ infers that the sensor $\mathcal{S}^{(I)}$ in the I-th building zone is *possibly* faulty, because it cannot conclude if only the sensor fault

f_1^e has occurred, provoking the violation of $\mathcal{E}^{(I)}$ or both f_1^e and $f^{(I)}$ have occurred. On the other hand, in the second fault scenario, where the sensors of all building zones become faulty, but the temperature sensor of the cooling coil is healthy, the agent $\mathcal{M}^{(I)}$ not only detects the presence of sensor faults but also isolates the sensor fault in the I-th building zone. This is realized in conjunction with the decision of the module $\mathcal{M}_1^e$ (Fig. 2a). In other words, all monitoring agents $\mathcal{M}^{(1)} - \mathcal{M}^{(8)}$ can isolate in a distributed manner the consecutive occurrence of multiple sensor faults in all zones (Fig.3c-3j). Particularly, when the agent $\mathcal{M}^{(I)}$ detects sensor faults, the observed pattern equals to $D^{(I)}(t) = [1, 0]^\top$, which is consistent with the first column of the sensor fault signature matrix $F^{(I)}$ shown in Table II, leading to the diagnosis set $\mathcal{D}^{(I)}(t) = \{f^{(I)}\}$. By comparing the simulation results illustrated in Fig. 3c-3j to the simulation results in Fig. 2c-2j, it can be stated that the effects of the propagated sensor fault f_1^e on the residuals and adaptive thresholds of $\mathcal{M}^{(1)} - \mathcal{M}^{(8)}$ are much lower than the effects of the local sensor faults. Therefore, in the first sensor fault scenario, we may infer that the occurrence of the local fault $f^{(I)}$ is more likely to have provoked the violation of $\mathcal{E}^{(I)}$, $I \in \{3, 4, 5, 6\}$, than the occurrence of the propagated sensor fault f_1^e, and characterize the sensor $\mathcal{S}^{(I)}$ as faulty for all $I \in \{3, 4, 5, 6\}$.

The effects of sensor faults on the actual temperature of the cooling coil, chilled water tank and all building zones, as well as the temperature estimations derived by the modules $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and agent $\mathcal{M}^{(I)}$, $I \in \{1, \ldots, 8\}$ can be observed in Fig. 4 and 5. According to Fig. 4a, when the temperature sensor of the cooling coil becomes faulty, the backsteping controller perceives the positive fault variation in the sensor output as an increase in the temperature and generates chilled water flow rate aiming at decreasing the actual temperature of the cooling coil. Also, due to this sensor fault the estimation of the temperature in the cooling coil is 'faulty', i.e. different from the actual temperature. When the temperature sensor of the water tank becomes also faulty, based on Fig. 4a and 4b, the actual temperature in the chilled water tank is less influenced compared to the actual temperature of the cooling coil, while its estimation deviates from the actual temperature less than the estimation of the temperature in the cooling coil. The occurrence of the sensor faults in the cooling coil and the chilled water tank is not observable in the actual temperature of the zones and their estimations provided by the agents $\mathcal{M}^{(1)} - \mathcal{M}^{(8)}$. As expected, the actual temperature of the building zones and their estimations are directly affected by faults in their temperature sensors (Fig. 4e-4h and Fig. 5c-5j). Particularly, due to the positive variation in the output of the sensor in the I-th zone $I \in \{1, \ldots, 8\}$, the feedback linearization controller generates air flow rate (control input) aiming at decreasing the temperature in the I-th zone. Also, in the second scenario, it can be observed that the actual temperatures of the cooling coil and chilled water tank are not affected considerably by the occurrence of multiple sensor faults in all building zones (Fig. 5a and 5b). By comparing the simulation results illustrated in Fig. 5c-5j to the simulation results in Fig. 4c-4j, we can observe that the effects of the propagated sensor fault f_1^e on the the actual temperature of all
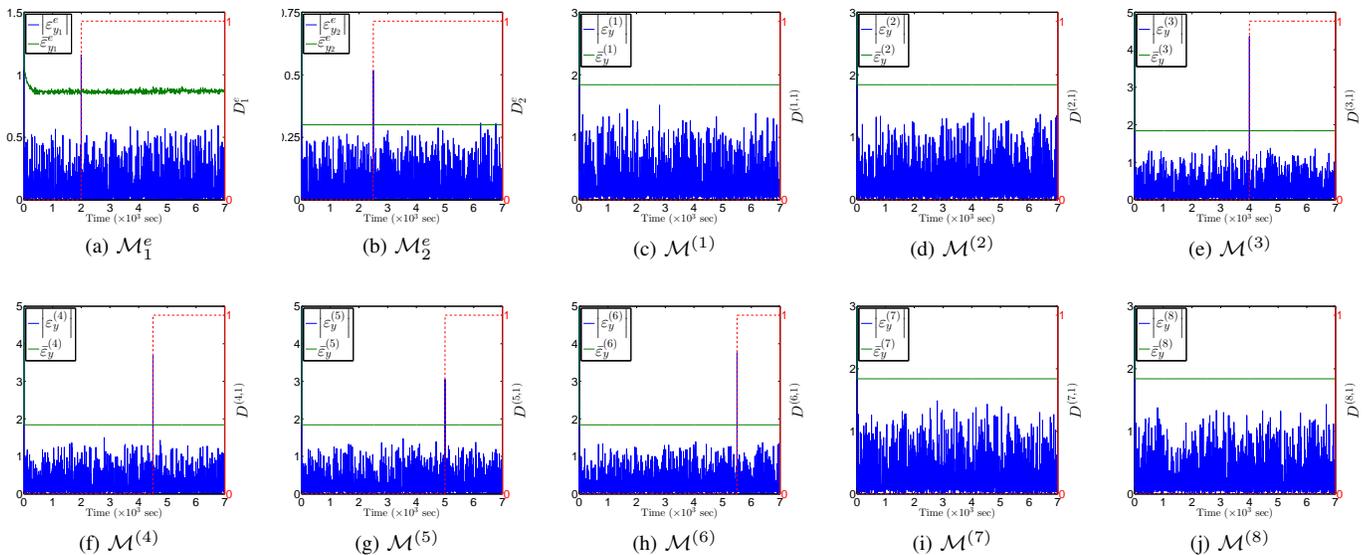
Fig. 2. Decision making-process of $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \ldots, 8\}$ for isolating multiple sensor faults that affect the electromechanical subsystem and building zones 3,4,5,6 consecutively. Every subfigure presents the temporal evolution of the magnitude of the residual (blue line) and the adaptive threshold (green line), as well as the boolean decision function (red dashed line).
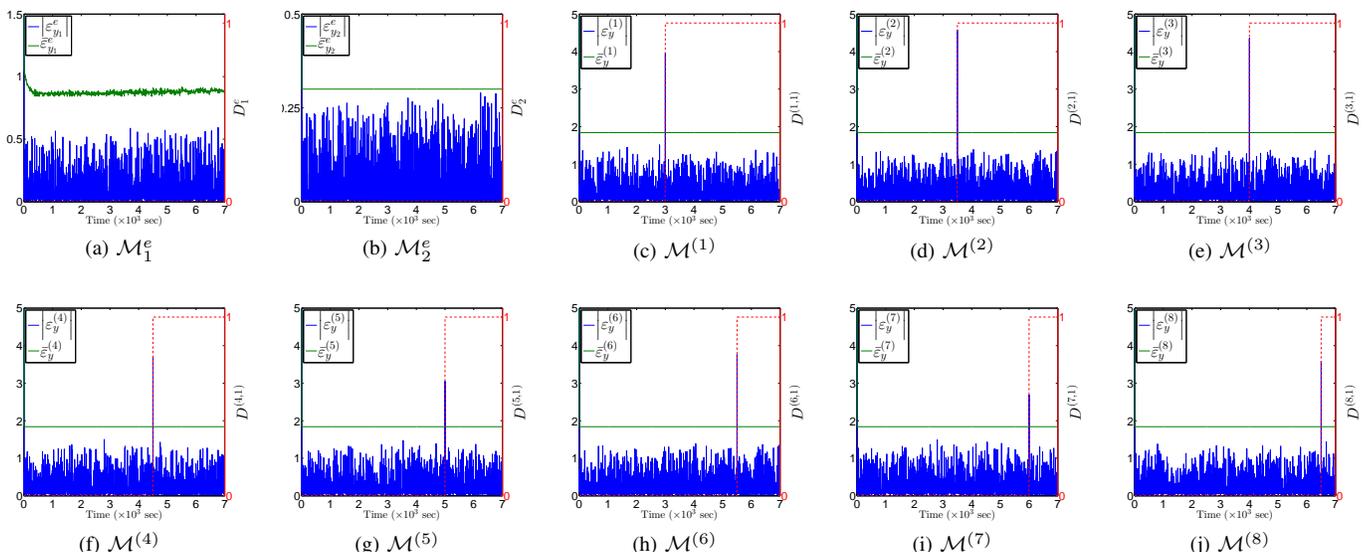

Fig. 3. Decision making-process of $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \ldots, 8\}$ for isolating multiple sensor faults that affect all building zones consecutively. Every subfigure presents the temporal evolution of the magnitude of the residual (blue line) and the adaptive threshold (green line), as well as the boolean decision function (red dashed line).

zones and the estimated temperature provided by $\mathcal{M}^{(1)} - \mathcal{M}^{(8)}$ are much lower than the effects of the local sensor faults.

VI. CONCLUSION

In this paper, we presented a model-based, distributed architecture for multiple sensor fault detection and isolation (SFDI) in a multi-zone HVAC system. The HVAC system was modeled as a set of interconnected subsystems. For each subsystem, we designed a local sensor fault diagnosis (LSFD) agent, where every agent was dedicated to each of the interconnected subsystems. The distributed isolation of multiple sensor faults was conducted by combining the decisions of the LSFD

agents and applying a reasoning-based decision logic. The performance of the proposed methodology was analyzed with respect to sensor fault detectability and multiple sensor fault isolability, characterizing the class of detectable and isolable sensor faults. The proposed SFDI technique may contribute to the reduction of energy consumption in large-scale buildings, as well as provide a procedure for the condition-base maintenance, thus reducing unnecessary maintenance work. Moreover the distributed deployment of the LSFD agents enhances the reliability with respect to security threats, while it is scalable for handling additional building zones in large-scale buildings. Simulation results illustrated the effectiveness of the
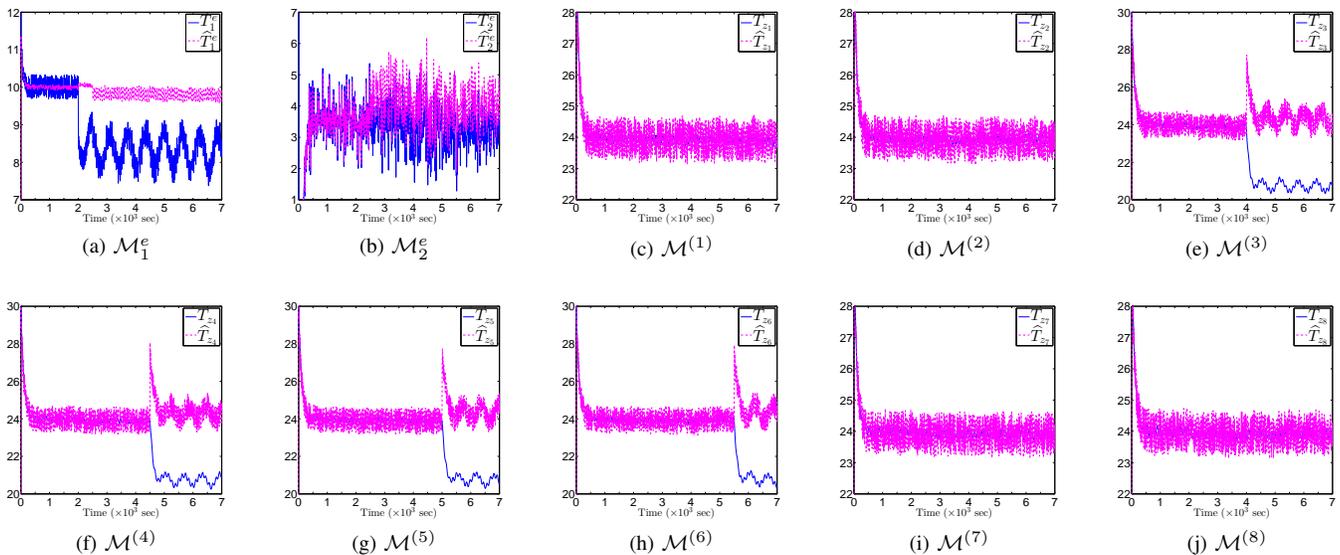
Fig. 4. Temperature estimation models (magenta dashed line) of $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \ldots, 8\}$ compared to actual temperatures (blue solid line) of the cooling coil, chilled water tank and building zones, under healthy conditions and consecutive occurrence of sensor faults in the electromechanical subsystem and building zones 3,4,5,6.
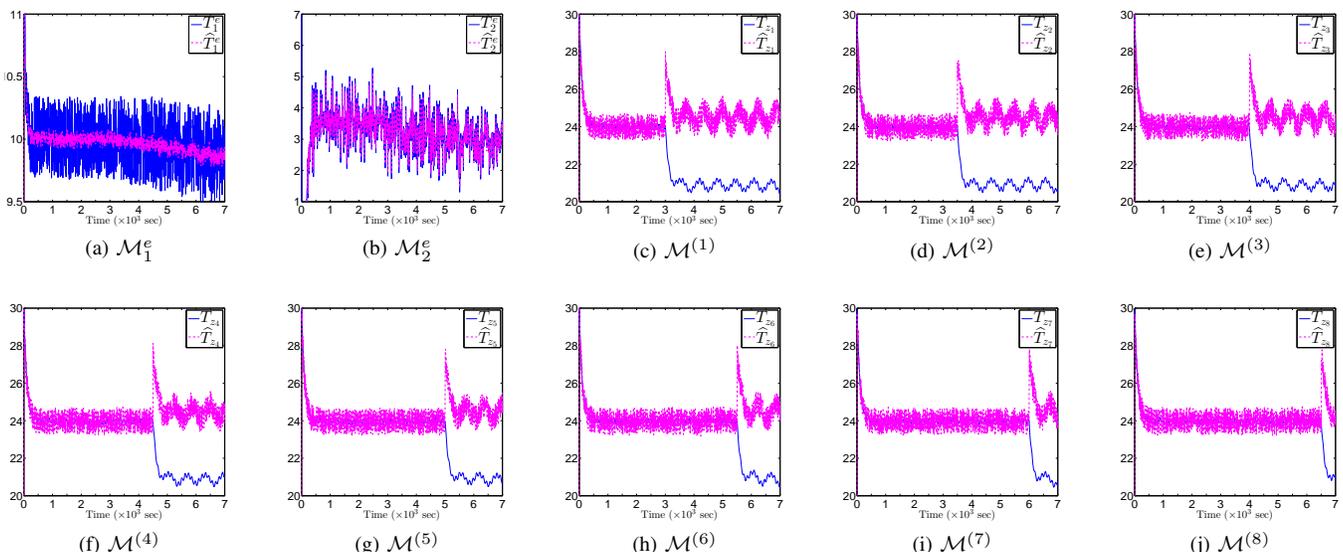

Fig. 5. Temperature estimation models (magenta dashed line) of $\mathcal{M}_1^e$, $\mathcal{M}_2^e$ and $\mathcal{M}^{(I)}$, $I \in \{1, \ldots, 8\}$ compared to actual temperatures (blue solid line) of the cooling coil, chilled water tank and building zones, under healthy conditions and consecutive occurrence of sensor faults in all building zones.

proposed distributed SFDI methodology in isolating multiple sensor faults in an eight-zone HVAC system.

Future research work will involve the integration of the proposed distributed methodology with other techniques for diagnosing both sensor and actuator faults in the HVAC system, aiming at resolving the problem of multiple sensor and actuator fault isolation, as well as the propagation of the sensor and actuator fault effects.

REFERENCES

[1] P. J. Antsaklis, B. Goodwine, V. Gupta, M. J. McCourt, Y. Wang, P. Wu, M. Xia, H. Yu, and F. Zhu, "Control of cyberphysical systems using passivity and dissipativity based methods," *European Journal of Control*, vol. 19, no. 5, pp. 379 – 388, 2013.
[2] K. W. J. Wong, "Development of selection evaluation and system intelligence analytic models for the intelligent building control systems," Ph.D. dissertation, The Hong Kong Polytechnic University, 2007.
[3] J. E. Braun, "Intelligent building systems-past, present, and future," in *Proceedings of 2007 American Control Conference*, 2007, pp. 4374–4381.
[4] R. Isermann, *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance.* Springer Verlag, 2006.

[5] J. Chen and R. Patton, *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic Publishers, 1999.

[6] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "A rule-based fault detection method for air handling units," *Energy and Buildings*, vol. 38, no. 12, pp. 1485–1492, 2006.

[7] B. T. Thumati, M. A. Feinstein, J. W. Fonda, A. Turnbull, F. J. Weaver, M. E. Calkins, and S. Jagannathan, "An online model-based fault diagnosis scheme for HVAC systems," in *Proceedings of IEEE International Conference on Control Applications (CCA)*, 2011, pp. 70–75.

[8] J. Liang and R. Du, "Model-based fault detection and diagnosis of HVAC systems using support vector machine method," *International Journal of Refrigeration*, vol. 30, no. 6, pp. 1104–1114, 2007.

[9] D. A. Veronica, "Detecting cooling coil fouling automatically –Part 2: Results using a multilayer perceptron," *HVAC&R Research*, vol. 16, no. 5, pp. 599–615, 2010.

[10] J. Hyvärinen and S. Kärki, *Building Optimization and Fault Diagnosis Source Book*. Technical Research Centre of Finland, VTT Building Technology, 1996.

[11] E. Miller-Hooks and T. Krauthammer, "An intelligent evacuation, rescue and recovery concept," *Fire Technology*, vol. 43, no. 2, pp. 107–122, 2007.

[12] J. Zhang, C. Mohan, P. Varshney, C. Isik, K. Mehrotra, S. Wang, Z. Gao, and R. Rajagopalan, "Intelligent control of building environmental systems for optimal evacuation planning," in *Proceedings of International Conference on Indoor Air Quality Problems and Engineering Solutions*, 2003, pp. 1–20.

[13] B. Fan, Z. Du, X. Jin, X. Yang, and Y. Guo, "A hybrid FDD strategy for local system of AHU based on artificial neural network and wavelet analysis," *Building and Environment*, vol. 45, no. 12, pp. 2698–2708, 2010.

[14] X. Xu, F. Xiao, and S. Wang, "Enhanced chiller sensor fault detection, diagnosis and estimation using wavelet analysis and principal component analysis methods," *Applied Thermal Engineering*, vol. 28, no. 2, pp. 226–237, 2008.

[15] Z. Du and X. Jin, "Detection and diagnosis for sensor fault in HVAC systems," *Energy Conversion and Management*, vol. 48, no. 3, pp. 693–702, 2007.

[16] S. M. Namburu, M. S. Azam, J. Luo, K. Choi, and K. R. Pattipati, "Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers," *IEEE Transactions on Automation Science and Engineering*, vol. 4, no. 3, pp. 469–473, 2007.

[17] S. Wang and J. Qin, "Sensor fault detection and validation of VAV terminals in air conditioning systems," *Energy Conversion and Management*, vol. 46, no. 15, pp. 2482–2500, 2005.

[18] S. Wang, Q. Zhou, and F. Xiao, "A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults," *Energy and Buildings*, vol. 42, no. 4, pp. 477–490, 2010.

[19] S. Katipamula and M. R. Brambley, "Methods for fault detection, diagnostics, and prognostics for building systems-A review, Part I," *HVAC&R Research*, vol. 11, no. 1, pp. 3–25, 2005.

[20] M.-L. Chiang and L.-C. Fu, "Hybrid system based adaptive control for the nonlinear HVAC system," in *Proceedings of 2006 American Control Conference*, 2006, pp. 5324–5329.

[21] M. Zaheer-Uddin, R. V. Patel, and S. A. Al-Assadi, "Design of decentralized robust controllers for multizone space heating systems," *IEEE Transactions on Control Systems Technology*, vol. 1, no. 4, pp. 246–261, 1993.

[22] M. Zaheer-Uddin, "Temperature control of multizone indoor spaces based on forecast and actual loads," *Building and Environment*, vol. 29, no. 4, pp. 485–493, 1994.

[23] A. Thosar, A. Patra, and S. Bhattacharyya, "Feedback linearization based control of a variable air volume air conditioning system for cooling applications," *ISA Transactions*, vol. 47, no. 3, pp. 339–349, 2008.

[24] W.-Y. Lee, J. M. House, and N.-H. Kyong, "Subsystem level fault diagnosis of a building's air-handling unit using general regression neural networks," *Applied Energy*, vol. 77, no. 2, pp. 153–170, 2004.

[25] S. Wang and J.-B. Wang, "Law-based sensor fault diagnosis and validation for building air-conditioning systems," *HVAC&R Research*, vol. 5, no. 4, pp. 353–380, 1999.

[26] ——, "Robust sensor fault diagnosis and validation in HVAC systems," *Transactions of the Institute of Measurement and Control*, vol. 24, no. 3, pp. 231–262, 2002.

[27] R. Escobar, C. Astorga-Zaragoza, A. Téllez-Anguiano, D. Juárez-Romero, J. Hernández, and G. Guerrero-Ramírez, "Sensor fault detection and isolation via high-gain observers: Application to a double-pipe heat exchanger," *ISA Transactions*, vol. 50, pp. 480–486, 2011.

[28] X. Yan and C. Edwards, "Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer," *International Journal of control*, vol. 81, no. 4, pp. 591–606, 2008.

[29] J. Weimer, J. Araujo, M. Amoozadeh, S. Ahmadi, H. Sandberg, and K. H. Johansson, "Parameter-invariant actuator fault diagnostics in cyber-physical systems with application to building automation," in *Control of Cyber-Physical Systems*, ser. Lecture Notes in Control and Information Sciences, 2013, vol. 449, pp. 179–196.

[30] R. M. G. Ferrari, T. Parisini, and M. M. Polycarpou, "Distributed fault detection and isolation of large-scale discrete-time nonlinear systems: An adaptive approximation approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 2, pp. 275–290, 2012.

[31] C. Keliris, M. Polycarpou, and T. Parisini, "A distributed fault detection filtering approach for a class of interconnected continuous-time nonlinear systems," *IEEE Transactions on Automatic Control,*, vol. 58, no. 8, pp. 2032–2047, 2013.

[32] Q. Zhang and X. Zhang, "Distributed sensor fault diagnosis in a class of interconnected nonlinear uncertain systems," *Annual Reviews in Control*, vol. 37, pp. 170 – 179, 2013.

[33] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Multiple sensor fault detection and isolation for large-scale interconnected nonlinear systems," in *Proceedings of European Control Conference*, Zurich, Switzerland, 2013, pp. 1952–1957.

[34] ——, "A distributed detection and isolation scheme for multiple sensor faults in interconnected nonlinear systems," in *52nd Conference on Decision and Control (CDC2013)*, Florence, Italy, 2013, pp. 4991–4996.

[35] V. Reppa, P. Papadopoulos, M. M. Polycarpou, and C. G. Panayiotou, "Distributed detection and isolation of sensor faults in HVAC systems," in *Proceedings of Mediterranean Conference on Control and Automation (MED)*, 2013, pp. 401–406.

[36] A. Talukdar and A. Patra, "Dynamic model-based fault tolerant control of variable air volume air conditioning system," *HVAC&R Research*, vol. 16, no. 2, pp. 233–254, 2010.

[37] M. Polycarpou and A. Trunov, "Learning approach to nonlinear fault diagnosis: detectability analysis," *IEEE Transactions on Automatic Control*, vol. 45, pp. 806–812, 2000.

[38] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Distributed sensor fault detection and isolation for nonlinear uncertain systems," in *8th IFAC SAFEPROCESS*, Mexico City, Mexico, 2012, pp. 1077–1082.

[39] W. Chen and J. Li, "Decentralized output-feedback neural control for systems with unknown interconnections," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, vol. 38, no. 1, pp. 258–266, 2008.

[40] P. Frank, "Handling modelling uncertainty in fault detection and isolation systems," *Journal of Control Engineering and Applied Informatics*, vol. 4, no. 4, pp. 29–46, 2002.

[41] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Springer Verlag, 2003.

[42] M. Cordier, P. Dague, F. Lévy, J. Montmain, M. Staroswiecki, and L. Travé-Massuyès, "Conflicts versus analytical redundancy relations: a comparative analysis of the model based diagnosis approach from the artificial intelligence and automatic control perspectives," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*, vol. 34, no. 5, pp. 2163–2177, 2004.

[43] V. Puig, A. Stancu, and J. Quevedo, "Robust fault isolation using nonlinear interval observers: the DAMADICS benchmark case study," in *16th IFAC World Congress*, 2005, pp. 1850–1855.

[44] M. Du, J. Scott, and P. Mhaskar, "Actuator and sensor fault isolation of nonlinear process systems," *Chemical Engineering Science*, vol. 104, pp. 294–303, 2013.

[45] S. Klinkhieo, R. J. Patton, and C. Kambhampati, "Robust FDI for FTC coordination in a distributed network system," in *16th IFAC World Congress*, Seoul, Korea, 2008, pp. 13 551–13 556.

[46] I. Samy, I. Postlethwaite, and D. Gu, "Survey and application of sensor fault detection and isolation schemes," *Control Engineering Practice*, vol. 19, pp. 658–674, 2011.

[47] H. K. Khalil, *Nonlinear Systems*. Prentice Hall, 2002.

[48] J. Farrell and M. Polycarpou, *Adaptive Approximation Based Control*. Wiley-Interscience, 2006.