Distributed Adaptive Estimation Scheme for Isolation of Sensor Faults in Multi-zone HVAC Systems

Panayiotis M. Papadopoulos*, Vasso Reppa**, Marios M. Polycarpou*, and Christos G. Panayiotou*

* KIOS Research Center for Intelligent Systems and Networks, Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, 1678, Cyprus (e-mail: {papadopoulos.panagiotis,mpolycar,christosp}@ucy.ac.cy)
** Department of Automatic Control, Supélec, Gif-sur-Yvette, 91192, France (e-mail: vasiliki.reppa@supelec.fr)

Abstract: This paper presents a model-based distributed scheme with emphasis on the isolation of sensor faults in multi-zone heating, ventilating and air-conditioning (HVAC) systems. A bank of local sensor fault detection and isolation agents are designed to diagnose sensor faults in a HVAC system, modeled as a set of interconnected, nonlinear subsystems. Each agent consists of the local sensor fault detection and adaptive estimation scheme for isolation of sensor faults. Detection and isolation signals are generated based on analytical redundancy relations (ARRs). These signals are provided to a local decision logic in order to distinguish between local and propagated sensor faults. Simulation results are used to illustrate the effectiveness of the proposed methodology applied to a four-zone HVAC system.

Keywords: distributed fault diagnosis, fault detection, fault isolation, sensor fault, HVAC system, smart buildings

1. INTRODUCTION

In recent years, there has been a high demand for energy efficiency, especially in large-scale systems such as transportation systems, power systems, smart buildings and many more. From a modeling perspective, a common characteristic of these systems is that they consist of various interconnected physical subsystems, which are monitored and controlled using a large number of sensors and/or sensor networks. The most energy consuming subsystem in a smart building is the heating, ventilating and air-conditioning (HVAC) system, which based on [Pérez-Lombard et al. (2008)], uses 50% of building consumption and 20% of total consumption in the USA. However, the smooth and healthy operation of the HVAC system is of significant importance in large-scale buildings for maintaining high indoor air quality for the occupants and sensitive equipment [ASHRAE (1996)], [Balaras et al. (2007)]. Furthermore, in safety-critical situations such as a fire or a contaminant release, the HVAC system must work properly in order to retain the applicability of control strategies executed by the Building Management System (BMS) for protecting the occupants [Haagras et al. (2003)].

A HVAC is a large-scale, complex system with many interconnected subsystems comprised of a several electrical and mechanical components for heating (i.e., boilers, heating coils, heat pump), cooling (i.e., cooling towers, chillers, cooling coils) and ventilating (i.e., fans, supply/return ducts, mixing boxes), as well as a large number of sensors, which may be located in different zones. It is inevitable that any of the HVAC components may be affected by faults, including: (i) actuator/process faults (i.e., fouled heat exchangers, stuck dampers, leaking valves, broken fans), (ii) sensor faults (i.e., drifting, stuck-at-zero), (iii) communication faults (i.e., wire break). In practice, it is hard to determine the presence and/or type of fault that has occurred unless its impact is either noticed by the occupants or technical staff, or by the high power consumption due to performance reduction. According to [Yu et al. (2002)], it is difficult to determine the location of the fault due to propagation through the physical interconnections between the subsystems of the HVAC system. Moreover, the occurrence of multiple faults makes the isolation procedure even more difficult.

Fault diagnosis methods can be classified into data-driven and analytical redundancy-based approaches [Isermann (2006)], [Gertler (1998)]. Data-driven methods use a classifier to identify faults, typically without taking into consideration the model of the system. Various of model-free fault diagnosis methods developed for HVAC systems using statistical methods including: neural networks approaches [Lee et al. (2004)], support vector machines [Liang and Du (2007)], principal component
analysis [Wang and Xiao (2004)], [Du and Jin (2007)], rule based methods [Schein et al. (2006)], and fuzzy approaches [Dexter and Ngo (2001)]. On the contrary, analytical redundancy-based fault diagnosis approaches rely on a qualitative or quantitative mathematical description of a system. Most of the analytical redundancy-based fault diagnosis methods for HVAC systems are focused either on local component area of the system [Wang et al. (2010)], [Schein et al. (2006)], [Bahel et al. (1995)] or on centralized approaches [Hou et al. (2006)], [Kumar and Kar (2013)]. The distributed fault detection and diagnosis approach in HVAC has not been investigated extensively [Repa et al. (2013, 2014)] due to large-scale of HVAC systems a distributed fault detection and diagnosis approach can facilitate the procedure of detection and isolation in terms of computational complexity and ensures physical redundancy.

The contribution of this paper is twofold. First, the proposed method is designed based on a multi-zone HVAC model with interconnected zones (i.e., the dynamics of a zone in the building are affected by the dynamics of their neighboring zones and the storage tank), making the implementation of a distributed sensor fault detection and isolation scheme more challenging, since the effect of a sensor fault can propagate to neighboring zones through a possibly distributed control scheme. Secondly, the proposed method is based on an adaptive estimation scheme, which differentiates the detection and isolation tasks. Taking into account the HVAC system topology, a bank of local sensor fault diagnosis agents are designed to detect and isolate single and multiple sensor faults using transmitted signals from their neighboring agents. Each local sensor fault diagnosis agent consists of a local sensor fault detection scheme that aims at detecting the occurrence of sensor faults in the monitored subsystem and its neighboring interconnected subsystems and a local adaptive estimation scheme designed to infer the occurrence of sensor faults that affect only its monitored subsystem (isolate local sensor faults) or infer also the propagation of sensor faults affecting subsystems interconnected to its monitored subsystem. For both schemes, analytical redundancy relations are designed to generate the local detection and isolation signals. A decision logic uses the local detection and isolation signals to determine the location of the sensor faults.

The paper is organized as follows. Section 2 provides a description of the multi-zone HVAC system used for the purposes of the distributed sensor fault detection and isolation methodology. In Section 3 the design of the distributed sensor fault detection and isolation architecture is described. The application of proposed methodology to a four-zone HVAC system is presented in Section 4. Some concluding remarks are given in Section 5.

2. MULTI-ZONE HVAC SYSTEM DESCRIPTION

This section presents the multi-zone HVAC model proposed by [Zaheer-Uddin (1994)], [Zaheer-Uddin and Patel (1995)]. Figure 1 describes the N-zone HVAC system with heating operation. The same structure of the HVAC system can be used also for cooling operation by replacing the heat pump with a chiller. The heat transferred from one zone to another is characterized by an inter-zone thermal flow coefficient which determines the amount of energy in the form of heat transmitted from the zone with the highest temperature in the zone with the lowest temperature. The water temperature in storage tank can be described by thermal-mass balance equation; i.e.,

\[
\frac{dT_{st}(t)}{dt} = \frac{U_{s,t,\text{max}}}{C_{st}} P_s(T_{st}(t)) u_{st}(t) + \frac{a_{st}}{C_{st}} \sum_{i \in N} u_i(t) U_{i,\text{max}} (T_i(t) - T_{st}(t)) - \frac{a_{st}}{C_{st}} (T_{st}(t) - T_{pl}(t)) + \frac{a_{st}}{C_{st}} T_{st}(t)
\]

where \( T_{st}(t) \) is the temperature of the water in storage tank, \( T_i(t) \) is the i-th zone temperature with \( i \in N \), \( a_{st} \) is the plenum (duct) temperature, \( T_{pl} \) (°C) is the plenum (duct) temperature, \( T_{st} \) (°C) is the source heat temperature to the heat pump, \( P_s(T_{st}(t)) \) is the performance coefficient of heat pump, \( T_{st}(t) \) (°C) is an uncertain term in temperature dynamics of the storage tank, \( p = (T_{\text{max}} - 1) \) where \( T_{\text{max}} \) is the rated maximum value of \( P_s(T_{st}) \), and \( \Delta T_{\text{max}} \) (°C) is the maximum temperature difference for the heat pump. The constant \( C_{st} \) (kJ/K) is the heat capacity of the storage tank, \( U_{i,\text{max}} \) (kg/h) is the maximum mass flow rate of hot water through the coil placed at \( i \)-th zone, and \( U_{st,\text{max}} \) (kJ/h) is the heat pump rated capacity. The coefficients \( a_{sz} \) and \( a_{st} \) (kJ/kg°C) refer to the effectiveness of the heating coil and the heat loss coefficient of storage tank from exterior surfaces, respectively. The parameter \( u_i \) is the mass flow rate of hot water flowing in the coil of \( i \)-th zone and \( u_{st} \) is the normalized energy in the heat pump. The \( i \)-th zone temperature dynamics can be presented by a first-order nonlinear differential equation as

\[
\frac{dT_i(t)}{dt} = \frac{U_{i,\text{max}} a_{sz}}{C_{zi}} (T_{st}(t) - T_{zi}(t)) u_i(t) - \frac{a_{sz}}{C_{zi}} (T_{zi}(t) - T_{amb}(t)) - \frac{h A_{zi}}{C_{zi}} (T_i(t) - T_{zi}(t)) - \frac{1}{C_{zi}} \sum_{j \in K_i} a_{zij} (T_{zi}(t) - T_{zj}(t)) + \frac{a_{sz}}{C_{zi}} T_{zi}(t),
\]

Fig. 1. Side view schematic representation of the N-zone HVAC system: supply water tubes (red dotted lines), return water tubes (blue solid lines).
where $T_{11}$ ($^\circ$C) is the temperature of the surface node of the mass wall in the $i$-th zone, $T_{amb}$ ($^\circ$C) is the ambient temperature, $T_z$ ($^\circ$C) is an uncertain term in temperature dynamics of the $i$-th zone, $h$ (W/m² $^\circ$C) is the heat transfer coefficient, $A_{dzi}$ (m²) is the surface area of the mass wall, and $C_{dzi}$ (kJ/°C) is the air heat capacity of the $i$-th zone. The coefficient $a_z$ (kJ/h $^\circ$C) corresponds to the heat loss coefficient from $i$-th zone, and $a_z$ (kJ/h $^\circ$C) is the inter-zone heat loss coefficient between $i$-th and $j$-th zone with $j \in K_i$, $K_i = \{ j : a_{zj} \neq 0 \}$. It is noted that $K_i$ are the indices of zones interconnected with the $i$-th zone and the variables $T_{pl}(t)$, $T_{s}(t)$, $T_{amp}(t)$ and $T_{i1}(t)$ are usually measured, but, for the sake of simplicity in this work, they will be considered as known.

The equations (1)-(2) and (3) describe the subsystems $\Sigma^s$ and $\Sigma^{(i)}$, $i \in \mathcal{N}$, respectively. Assuming that water temperature at the storage tank and the source temperature under healthy and faulty conditions, satisfy $T_st(t) \leq \Delta T_{max} + T_{o}(t)$, the performance coefficient $P_s(t)$ is equal to the top term of (2) for all $t$, the subsystem $\Sigma^s$ can be represented by a nonlinear state space form as

$$
\Sigma^s: \dot{x}^s(t) = A^s x^s(t) + g^s(x^s(t),d^s(t))u^s(t) \\
+ h^s(x^s(t),x(t),u(t)) + \eta^s(d^s(t)) + r^s(t), \quad (4)
$$

where $x^s \equiv T_{st}, u^s \equiv u_{st}$ and $r^s(t) \equiv \frac{\eta^s}{C_{st}}T_{st}$ represent the state, control input and modeling uncertainty of subsystem $\Sigma^s$, respectively, $x \equiv [x^{(1)},\ldots,x^{(N)}] \equiv [T_{z1},\ldots,T_{zn}]$ is the interconnection vector, $u \equiv [u^{(1)},\ldots,u^{(N)}] \equiv [u_{11},\ldots,u_{nn}]$ is the input vector of the neighboring subsystems $\Sigma^{(1)}, \ldots, \Sigma^{(N)}$, and $d^s \equiv [d_1^s,d_2^s]^T \equiv [T_{pl},T_{o}]^T$ is uncontrollable but known exogenous input vector. The parameter $A^s$, which is a coefficient for the linear local dynamics of the subsystem $\Sigma^s$, is defined as $A^s = -\frac{a_{z}}{C_{st}}$. The terms $g^s(\cdot)u^s$ and $h^s(\cdot)$ describe the nonlinear local and interconnection dynamics of the subsystem $\Sigma^s$, respectively, defined as

$$
g^s(x^s,d^s) = \frac{U_{i,\text{max}}}{C_{st}}\left(1 + p\left(1 - \frac{x^s(t) - d^s(t)}{\Delta T_{\text{max}}}ight)\right),
$$

$$
h^s(x^s,t,x(t),u(t)) = \% \sum_{i \in \mathcal{N}} U_{i,\text{max}}(x^s(t) - x^s(t))u^s(t), \quad \text{and} \quad \eta^s(\cdot)
$$

represents the known disturbances of the subsystem $\Sigma^s$ as $\eta^s(d^s) = \% \sum_{i \in \mathcal{N}} U_{i,\text{max}}d^s_i$.

Similarly, based on the differential equation (3), we obtain the state space representation of the subsystem $\Sigma^{(i)}$, $i \in \mathcal{N}$, i.e.,

$$
\Sigma^{(i)}: \dot{x}^{(i)}(t) = A^{(i)}x^{(i)}(t) + g^{(i)}(x^{(i)}(t),u^{(i)}(t)) \\
+ h^{(i)}(x^{(i)}(t),x^{(j)}(t),u^{(i)}(t)) + \eta^{(i)}(d^{(i)}(t)) + r^{(i)}(t), \quad (5)
$$

where $A^{(i)} = \frac{h_{A_{ dzi}}}{C_{zi}} - \frac{1}{C_{zi}} \sum_{j \in K_i} a_{zi,j}, x^{(i)} \equiv T_{zi}, u^{(i)} \equiv u_i$ and $r^{(i)} = \% \sum_{j \in K_i} U_{i,\text{max}}a_{zi,j}T_{zi}$ represent the state variable, control input and the modeling uncertainty of the subsystem $\Sigma^{(i)}$, $i \in \mathcal{N}$, respectively, where $N$ is the number of zones of the HVAC system. The terms $g^{(i)}(\cdot)u^{(i)}$ and $h^{(i)}(\cdot)$ denote the nonlinear local and interconnection dynamics of the subsystem $\Sigma^{(i)}$, respectively with $g^{(i)}(x^{(i)}(t),u^{(i)}(t)) = -\frac{U_{i,\text{max}}a_{zi,j}}{C_{zi}}x^{(i)}(t)u^{(i)}$, $h^{(i)}(x^{(i)},x^{(j)},u^{(i)}) = \% \sum_{j \in K_i} a_{zi,j}x^{(i)} + \frac{U_{i,\text{max}}a_{zi,j}}{C_{zi}}x^{(i)}u^{(i)}$ where $j \in K_i$ ($K_i$ is the set of indices of the zones that are interconnected with the $i$-th zone). The known disturbances of the subsystem $\Sigma^{(i)}$ are defined by the term $\eta^{(i)}(\cdot)$ as $\eta^{(i)}(d^{(i)}(t)) = \% \left(\frac{a_{zi,j}}{C_{zi}}d_{j}^{(i)} - \frac{h_{A_{dzi}}}{C_{zi}}d_{j}^{(i)}\right)$, where $d^{(i)} = \% \left[d_{1}^{(i)},d_{2}^{(i)}\right]^T \equiv [T_{amb},T_{i1}]^T$ is the uncontrollable but known exogenous input vector. The subsystems $\Sigma^s$ and $\Sigma^{(i)}$ are monitored and controlled using the sensors $S^s$ and $S^{(i)}$, respectively, which are characterized by the outputs $y^s \in R$ and $y^{(i)} \in R$; i.e.,

$$
S^s: \quad y^s(t) = x^s(t) + n^s(t) + f^s(t), \quad (6)
$$

$$
S^{(i)}: \quad y^{(i)}(t) = x^{(i)}(t) + n^{(i)}(t) + f^{(i)}(t), \quad (7)
$$

where $y^s$ and $y^{(i)}$ are the sensor measurements of water temperature in storage tank and air temperature in zone $i$, respectively, $n^s$ and $n^{(i)}$ represent the noise corrupting the sensor measurements $y^s$ and $y^{(i)}$, respectively, and $f^s$ and $f^{(i)}$ are the sensor faults. The system presented in (1)-(3) is controlled by $N + 1$ distributed nonlinear feedback controllers. Every controller uses local and neighboring sensor measurements, and control input signals of the neighboring interconnected subsystems for achieving its control objective that is tracking some desired reference signals for the temperature in every subsystem (i.e., $b^s$ and $b^{(i)}$ are the reference signals for the states $x^s$ and $x^{(i)}$, $i \in \mathcal{N}$, respectively).

### 3. DISTRIBUTED SENSOR FAULT DETECTION AND ISOLATION SCHEME

A set of distributed sensor fault detection and isolation agents denoted by $\mathcal{M}^s$ and $\mathcal{M}^{(i)}$ have been developed for the subsystems $\Sigma^s$ and $\Sigma^{(i)}$, $i \in \mathcal{N}$, respectively. Every agent contains a local sensor fault detection and a local sensor fault isolation scheme.

#### 3.1 Distributed Sensor Fault Detection

The decision of the agent $\mathcal{M}^s$ and $\mathcal{M}^{(i)}$ about the occurrence of sensor faults is obtained by checking the satisfaction of analytical redundancy relations of residuals and adaptive thresholds, as presented next.

**Residual Generation:**

For estimating the water temperature in the storage tank a nonlinear state estimation model is used by $\mathcal{M}^s$; i.e.,

$$
\hat{x}^s(t) = A^s\hat{x}^s(t) + g^s(y^s(t),d^s(t))u^s(t) \\
+ h^s(y^s(t),y(t),u(t)) + \eta^s(d^s(t)) + L^s(y^s(t) - \hat{x}^s(t)), \quad (8)
$$

where $\hat{x}^s$ is the estimation of $x^s$, $y \equiv [y^{(1)},\ldots,y^{(N)}]^T$ is a vector that contains the sensor measurements shared through agents $\mathcal{M}^{(i)}$, $i \in \mathcal{N}$ and $L^s$ is Luenberger estimation gain, such that $A_L^s \equiv A^s - L^s$ is a stable. The residual $\varepsilon_y^s$ is defined by

$$
\varepsilon_y^s(t) = y^s(t) - \hat{x}^s(t), \quad (9)
$$

where $y^s$ is the measurements of the sensor $S^s$. The residual $\varepsilon_y^s$ can be expressed also as
\[ e_y^*(t) = e^{A_y^* t} e_y^*(0) + n^*(t) + \int_0^t e^{A_y^*(t-\tau)} \left( r^*(\tau) + g^*(x^*(\tau), d^*(\tau)) \right) u^*(\tau) - h^*(y^*(\tau), u^*(\tau)) - L^* n^*(\tau) \, d\tau. \] 

(10)

Note that the residual \( e_y^* \) can be affected by sensor faults occurring in sensors \( S^* \) and \( S(i) \), \( i \in \mathcal{K}_i \).

For estimating the air temperature in zone \( i \), a nonlinear estimation model associated with \( M(i) \), \( i \in \mathcal{K}_i \) is generated by

\[ \hat{x}^i(t) = A^i \hat{x}^i(0) + n^i(t) + \int_0^t e^{A^i(t-\tau)} \left( r^i(\tau) - L^i n^i(\tau) \right) + g^i(y^i(\tau), u^i(\tau)) + h^i(x^i(\tau), y^i(\tau), u^i(\tau)) \, d\tau. \]

(11)

where \( \hat{x}^i \) is the estimation of \( x^i \) and \( L^i \) is Luenger estimation gain, such as \( A^i \equiv A(i) - L^i \) is a stable. The residual \( e_y^i \) is defined by

\[ e_y^i(t) = y^i(t) - \hat{x}^i(t), \]

(12)

where \( y^i \) is the measurements of the sensor \( S^i \), \( i \in \mathcal{N} \).

The residual \( e_y^j \), \( i \in \mathcal{N} \) can be also expressed as

\[ e_y^j(t) = e^{A^j t} e_y^j(0) + n^j(t) + \int_0^t e^{A^j(t-\tau)} \left( r^j(\tau) - L^j n^j(\tau) \right) + (g^j(x^j(\tau)) - g^j(y^j(\tau), u^j(\tau))) u^j(\tau) + h^j(x^j(\tau), y^j(\tau), u^j(\tau)) - h^j(y^j(\tau), u^j(\tau)) \, d\tau. \]

(13)

where \( n^j(0) \leq \bar{n}^j \), \( |r^j(\tau)| \leq \bar{r}^j \) and \( |u^j(\tau)| \leq \bar{u}^j \).

Note that the residual \( e_y^j \) can be affected by sensor faults occurring in sensors \( S^j \), \( j \in \mathcal{K}_j \).

**Computation of Adaptive Thresholds:**

The adaptive thresholds bound the residual under healthy conditions taking into account the following assumptions:

**Assumption 1:** The input variables \( u^* \) and \( u^j \), and state variables \( x^* \) and \( x^j \), \( i \in \mathcal{N} \) have to be bounded in the compact spaces \( \mathcal{U} \) and \( \mathcal{X} \), respectively.

**Assumption 2:** The noise corrupting the measurements and the modeling uncertainty are unknown functions but uniformly bounded by positive constants \( \bar{n}^* \), \( \bar{n}^j \), \( \bar{r}^* \) and \( \bar{r}^j \) which are known for all \( i \), such that \( |n^*(t)| \leq \bar{n}^* \), \( |n^j(t)| \leq \bar{n}^j \), \( |r^*(t)| \leq \bar{r}^* \) and \( |r^j(t)| \leq \bar{r}^j \), respectively.

The adaptive threshold \( e_y^* \) for the agent \( M^* \) is designed based on (10) and assuming healthy conditions (i.e., \( f^* = 0 \), \( f^j = 0 \), \( i \in \mathcal{N} \)) such that \( |e^*(t)| \leq \bar{e}^* \). The adaptive threshold \( \bar{e}_y^* \) is defined as

\[ \bar{e}_y^*(t) = \rho^* e^{-\xi^* t} \bar{x}^* + \bar{n}^* + \int_0^t \rho^* e^{-\xi^* (t-\tau)} \left( |L^*| \bar{n}^* \right) + \bar{r}^* + \frac{\bar{M}}{\bar{C}_t} \sum_{i \in \mathcal{N}} U_{i, \text{max}} \left( \bar{n}^i + \bar{r}^i \right) u^i(\tau) \, d\tau, \]

(14)

where \( |x^*(0)| \leq \bar{x}^* \) and \( |e^{A_y^* t}| \leq \rho^* e^{-\xi^* t} \) for all \( t \). An analytical redundancy relation is constructed based on the residual \( e_y^* \) and adaptive threshold \( \bar{e}_y^* \) developed in (9) and (14), respectively, such that

\[ \mathcal{E}^* : |e_y^*(t)| \leq \bar{e}_y^*(t). \]

(15)

If \( \mathcal{E}^* \) is not satisfied, the boolean decision signal \( D^* \) is defined as

\[ D^*(t) = \begin{cases} 1, & t \geq T_d^* \\ 0, & t < T_d^* \end{cases}, \]

(16)

where \( T_d^* \) is the detection time, such as \( T_d^* = \inf \{ t \geq 0 : |e_y^*(t)| > \bar{e}_y^*(t) \} \). Additionally, the adaptive threshold \( \bar{e}_y^j \) for the agent \( M(j) \) is designed based on (13) and assuming healthy conditions (i.e., \( f^* = 0 \), \( f^j = 0 \), \( i \in \mathcal{N} \)) such that \( |e_y^j(t)| \leq \bar{e}_y^j(t) \). Thus, the adaptive threshold \( \bar{e}_y^j \) is defined by

\[ \bar{e}_y^j(t) = \rho^j e^{-\xi^j t} \bar{x}^j + \bar{n}^j + \int_0^t \rho^j e^{-\xi^j (t-\tau)} \left( |L^j| \bar{n}^j \right) + \bar{r}^j + \frac{\bar{M}}{\bar{C}_t} \sum_{i \in \mathcal{N}} U_{i, \text{max}} \left( \bar{n}^i + \bar{r}^i \right) u^i(\tau) \, d\tau, \]

(17)

where \( |x^j(0)| \leq \bar{x}^j \) and \( |e^{A_y^j t}| \leq \rho^j e^{-\xi^j t} \) for all \( t \). Similarly, an analytical redundancy relation is formulated based on the residual \( e_y^j \) and adaptive threshold \( \bar{e}_y^j \) developed in (13) and (17), respectively, such that

\[ \mathcal{E}^j : |e_y^j(t)| \leq \bar{e}_y^j(t). \]

(18)

If \( \mathcal{E}^j \) is not satisfied, the boolean decision signal \( D^j \) is defined as

\[ D^j(t) = \begin{cases} 1, & t \geq T_d^j \\ 0, & t < T_d^j \end{cases}, \]

(19)

where \( T_d^j \) is the detection time, such as \( T_d^j = \inf \{ t \geq 0 : |e_y^j(t)| > \bar{e}_y^j(t) \} \), \( i \in \mathcal{N} \).

**3.2 Distributed Sensor Fault Isolation**

**Adaptive Estimation Scheme:**

When \( D^* = 1 \) and \( D^j = 1 \), the corresponding isolation schemes are activated. Every local isolation scheme is designed assuming that only a local sensor fault has occurred and the goal is to verify this hypothesis. The adaptive estimation scheme associated with \( M^* \) is described by

\[ \dot{x}_q^*(t) = A_x^* x_q^*(t) + g^*(y_q^*(t) - \bar{y}_q^*(t), d^*(t)) u^*(t) + h^*(y_q^*(t) - \bar{y}_q^*(t), y_q^*(t), u^*(t)) + \eta^*(d^*(t)) \]

\[ + L^* (y_q^*(t) - \bar{y}_q^*(t)) + \Omega_q^* (\bar{y}_q^*(t) - \bar{y}_q^*(t)) + \bar{x}_q^*(T_q^*) = 0 \]

\[ \dot{\bar{y}}_q^*(t) = A_{\bar{y}}^* \bar{y}_q^*(t) - L^* + \frac{\bar{M}}{\bar{C}_t} \sum_{i \in \mathcal{N}} U_{i, \text{max}} \left( \bar{n}^i + \bar{r}^i \right) u^i(\tau) \]

\[ + \frac{\bar{M}}{\bar{C}_t} \sum_{i \in \mathcal{N}} U_{i, \text{max}} u^i(\tau), \Omega_q^*(T_q^*) = 0 \]

(20)
where $\dot{x}^*_q$ is the state estimation of the state $x^*$, $\dot{\theta}^*_q$ is the estimation of the fault \( f^* \), $\Omega^*_q$ is filtering term [Farrell and Polycarpou (2006)]. Note that $\epsilon^*_q$ is residual of the distributed adaptive estimation scheme of agent $M^*$.

Similarly, the adaptive estimation scheme associated with $M^*(i)$, $i \in \mathcal{K}$ is described by

$$
\dot{x}^*_q(t) = A^*(i)\dot{x}^*_q(t) + g^*(i)(y^*(i)(t) - \hat{y}^*_q(t))u(i)(t) + h^*(i)(y^*(i)(t), u(t)) + \Omega^*_q(t)\dot{\theta}^*_q(t), \quad \dot{\theta}^*_q(T_d^*(i)) = 0
$$

$$
\hat{\mathbf{y}}^*_q(t) = A^*_L(\hat{\mathbf{y}}^*_q(t) - \mathbf{I}^*(i) - \frac{U_i}{\epsilon_{s_{\max}}^{(i)}}u(i)(t), t), \quad \Omega^*_q(T_d^*(i)) = 0
$$

$$
\dot{\gamma}^*_q(t) = \dot{\gamma}^*_q(t) + \hat{\theta}^*_q(t)
$$

$$
\epsilon^*_q(t) = y^*(i)(t) - \hat{y}^*_q(t)
$$

$$
\hat{\theta}^*_q(t) = \gamma^*_q(\Omega^*_q(t) + 1)\epsilon^*_q(t),
$$

where $\dot{x}^*_q$ is the state estimation of the state $x^*$, $\dot{\theta}^*_q$ is the approximation of the magnitude of the fault function $f^*(i)$, $\Omega^*_q(t)$ is a filtering term and $\epsilon^*_q(t)$ is the residual of the distributed adaptive estimation scheme in agent $M^*(i)$, $i \in \mathcal{K}$.

**Computation of Adaptive Thresholds**

The adaptive threshold $\epsilon^*_y$ for distributed sensor fault isolation is defined as

$$
\epsilon^*_y(t) = \kappa^* e^{-\lambda^* t} |x^*_q| + L^* + |\Omega^*_q(t) + 1|\delta^* + \int^{t}_{t_1} \kappa^* e^{-\lambda^* (t-\tau)} \left[ L^* + \frac{U_{i_{\max}}(P_{\max}-1)}{\epsilon_{s_{\max}}^{(i)}} \hat{\eta}^*(\tau) \right] d\tau + \frac{\alpha^*}{\epsilon_{s_{\max}}^{(i)}} \sum_{j \in \mathcal{K}} U_{i_{\max}} \left( \hat{\eta}^*(\tau) \left| u(i)(\tau) \right| \right) d\tau,
$$

where $|x^*_q(T_d^*)| \leq \bar{x}^*_s$, $|e^{A^*_i}t| \leq \kappa^* e^{-\lambda^* t}$, and $|f^*(t) - \hat{\theta}^*_q(t)| \leq \delta^*$. The design parameters $\bar{x}^*_s$, $\kappa^*$, $\lambda^*$ and $\delta^*$ are positive constant values such that the above conditions are satisfied for all $t$.

An analytical redundancy relation is constructed based on the residual $\epsilon^*_y$ and adaptive threshold $\epsilon^*_y$ developed in (21) and (22), respectively, such that

$$
\mathcal{R}^*: \left| \epsilon^*_y(t) \right| \leq \bar{\epsilon}^*_y(t).
$$

If $\mathcal{R}^*$ is not satisfied, the boolean isolation signal $I^*$ for agent $M^*$ is defined as

$$
I^*(t) = \begin{cases} 1, & t \geq T^*_f \\ 0, & t < T^*_f \end{cases}
$$

where $T^*_f$ is the isolation time, such as $T^*_f \equiv \inf \{t \geq 0: \left| \epsilon^*_y(t) \right| > \bar{\epsilon}^*_y(t) \}$.

The adaptive threshold $\epsilon^*_y(i)$ for distributed sensor fault isolation is computed as

$$
\epsilon^*_y(i)(t) = \kappa(i) e^{-\lambda(i) t} |x^*_q(i)| + \bar{\epsilon}^*_y(i) + \left( |\Omega^*_q(i)(t) + 1| \delta^* + \left( \frac{U_{i_{\max}}(P_{\max}-1)}{\epsilon_{s_{\max}}^{(i)}} \hat{\eta}^*(\tau) \right) + \frac{\alpha^*}{\epsilon_{s_{\max}}^{(i)}} \sum_{j \in \mathcal{K}} U_{i_{\max}} \left( \hat{\eta}^*(\tau) \left| u(i)(\tau) \right| \right) d\tau, \right.
$$

where $|x^*_q(T_d^*)| \leq \bar{x}^*_s$, $|e^{A^*_i}t| \leq \kappa(i) e^{-\lambda(i) t}$, and $|f(i)(t) - \hat{\theta}^*_q(t)| \leq \delta(i)$. The design parameters $\bar{x}^*_s$, $\kappa(i)$, $\lambda(i)$ and $\delta(i)$ are positive constant values such that the above conditions are satisfied for all $t$.

Similarly, an analytical redundancy relation is formulated based on the residual $\epsilon^*_y(i)$ and adaptive threshold $\epsilon^*_y(i)$ developed in (21) and (25), respectively, such that

$$
\mathcal{R}^*: \left| \epsilon^*_y(i)(t) \right| \leq \bar{\epsilon}^*_y(i)(t).
$$

If $\mathcal{R}^*$ is not satisfied, the boolean isolation signal $I^*(i)$ for the agent $M^*(i)$ is defined as

$$
I^*(i)(t) = \begin{cases} 1, & t \geq T^*_f(i) \\ 0, & t < T^*_f(i) \end{cases}
$$

where $T^*_f(i)$ is the isolation time, such as $T^*_f(i) \equiv \inf \{t \geq 0: \left| \epsilon^*_y(i)(t) \right| > \bar{\epsilon}^*_y(i)(t) \}$, $i \in \mathcal{K}$.

**3.3 Distributed Sensor Fault Isolation Decision Logic**

When the agent $M^*$ (similarly for the agent $M^*(i)$) detects the presence of sensor faults, i.e., $D^*=1$ (similarly, if $D^*(i)=1$), then the isolation scheme is activated. As long as $I^*=0$, then the agent $M^*$ (correctly) infers that the sensor $S^*$ in the storage tank is faulty. If $I^*=1$, then the agent $M^*$ infers that the temperature sensor in at least one zone is faulty (propagated sensor faults) while the sensor in the storage tank is possibly faulty. As long as $I^*(i)=0$, the agent $M^*(i)$ infers that the temperature sensor in its monitored zone is faulty. If $I^*(i)=1$, the agent $M^*(i)$ infers that at least one temperature sensor in the zones that are interconnected with its monitored zone, or the storage tank is faulty (propagated sensor fault), while the temperature sensor in its monitored zone is characterized as possibly faulty. Under the assumption of a single sensor fault affecting one of the subsystems, if $I^*=1$ or $I^*(i)=1$, then the agents $M^*$ and $M^*(i)$ infer the occurrence of propagated faults only, excluding the occurrence of local sensor faults.

**4. SIMULATION RESULTS**

In this section, we illustrate the fault diagnosis methodology developed in Section 3 on a four-zone HVAC system, based on the model presented in (1)-(3), where $\mathcal{K} = \{1, 2, 3, 4\}$. As a result, based on (4)-(5), the following five subsystems are considered: $\Sigma^*, \Sigma^1, \Sigma^2, \Sigma^3, \Sigma^4$. The subsystem $\Sigma^*$ is interconnected with each $\Sigma^j(i), i \in \mathcal{K}$. The architectural arrangement of the four zones is presented in Figure 2(a), while the interconnections are depicted in Figure 2(b). Based on these interconnections, we defined the sets $K_1 = \{2, 3\}, K_2 = \{1, 3, 4\}, K_3 = \{1, 2, 4\}, K_4 = \{2, 3\}$. 

To simplify the diagram, we did not include $\Sigma^s$, which is connected to each of the four subsystems $\Sigma^{(1)}$, $\Sigma^{(2)}$, $\Sigma^{(3)}$, $\Sigma^{(4)}$, which represent the four zones. The parameters of the four-zone HVAC system are defined as follows: $a_{11}=a_{22}=740$, $a_{12}=a_{21}=540$, $a_{13}=a_{14}=a_{23}=a_{24}=a_{34}=500$, $a_{31}=12$, $a_{32}=6$, $C_{st}=837$, $C_{z1}=C_{z2}=370$, $C_{z3}=250$, $U_{1,max}=U_{3,max}=3700$, $U_{2,max}=U_{4,max}=2700$, $A_{d,1}=A_{d,2}=A_{d,3}=A_{d,4}=120$, $U_{s1,max}=27.36 \times 10^4$, $P_{max}=3.5$, $\Delta T_{max}=45$ and $h=8.29$. It is assumed that the exogenous inputs are constant and defined as follows: $T_0 = 5^\circ C$, $T_{pl} = 10^\circ C$, $T_{amb} = 5^\circ C$, $T_{11} = T_{31} = 10^\circ C$, $T_{21} = T_{31} = 12^\circ C$. The system is controlled by five distributed linearization controllers (i.e., four dedicated controllers for each subsystem taking into account the temperature dynamics of the four zones and one dedicated controller for the subsystem taking into account the temperature dynamics of the storage tank). The controller gains are selected as $k^s=6$, $k^{(i)}=4$, $i \in \mathcal{X}$. The desired values of the temperatures are set up as follows: $h^s=55^\circ C$ and $h^{(i)}=22^\circ C$, $i \in \mathcal{X}$. The modeling uncertainty in each subsystem $r^s = 5\%d^s sin(0.1t)$ and $r^i = 5\%d^{(i)} sin(0.1t)$, $i \in \mathcal{X}$. For simulation purposes, the noise corrupting the sensor output is defined as follows: $n^s = 3\%Y^s$ and $n^{(i)} = 3\%Y^{(i)}$, where $Y^s$ and $Y^{(i)}$ are the steady state value of sensor measurements $y^s$ and $y^{(i)}$, respectively, $i \in \mathcal{X}$, under healthy conditions. The design constants for the distributed sensor fault detection and isolation (DSFDI) agents are chosen as follows: $L^s=5$ and $L^{(i)}=4$, $p^s=1.3$, $p^{(i)}=1.3$, $\xi^s=10$, $\xi^{(i)}=6$, $\gamma^s=3$, $\gamma^{(i)}=3$, $\delta^s=8$, $\delta^{(i)}=3$, $\lambda^s=8$, $\lambda^{(i)}=8$, $\kappa^s=1$, $\kappa^{(i)}=1$. In the simulation experiment, we consider the following multiple sensor fault scenario: Two bias abrupt faults occur in sensors $S^s$ and $S^{(4)}$ at simulation time $t^s_f = t^{(4)}_f = 20 h$ such as $f^s(t) = 15\%Y^s(1 - e^{-50^2(t-20)})$ and $f^{(4)}(t) = 20\%Y^{(4)}(1 - e^{-i0^2(t-20)})$, while the form of sensor faults is presented in [Reppa et al. (2013, 2014)].

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig2.png}
\caption{Schematic representation of the four zones heated by the HVAC system: (a) architectural arrangement of zones, (b) internal temperature interactions (bidirectional arrows) between the subsystems ($\Sigma^{(1)}$, $\Sigma^{(2)}$, $\Sigma^{(3)}$, $\Sigma^{(4)}$).}
\end{figure}

In general, the simultaneous occurrence of more than one fault makes the fault isolation task more challenging, which is the reason for considering two simultaneous faults in the current scenario. Figures 3 and ?? present the simulation results for the multiple sensor scenario. In more detail, Figures 3(a)–3(e) show the residuals $\varepsilon^s_y$ and $\varepsilon^{(i)}_y$ (blue solid lines), the adaptive thresholds $\varepsilon^s_y$ and $\varepsilon^{(i)}_y$ (green solid lines), and the detection signals $D^s$ and $D^{(i)}$ (red dotted lines), $i \in \{1, 2, 3, 4\}$. As shown, at time instants $T^s_3 = 20.19 h$, $T^{(1)}_4 = 20.33 h$ and $T^{(4)}_d = 20 h$, the distributed sensor fault detection scheme of agents $M^s$, $M^{(1)}$ and $M^{(4)}$ detect, respectively, the sensor faults. Figures 3(f)–3(h) show the output residuals $\varepsilon^s_y$ and $\varepsilon^{(i)}_y$ (blue solid lines), the adaptive thresholds $\varepsilon^s_y$ and $\varepsilon^{(i)}_y$ (green solid lines), and the detection signals $I^s$ and $I^{(i)}$ (red dotted lines), $i \in \{1, 4\}$. At time instant $T^{(1)}_t = 21.46 h$, the isolation signal $I^{(1)}$ becomes 1, and the agent $M^{(1)}$ excludes the single occurrence of a fault affecting sensor $S^{(1)}$ and infers that at least one of the sensors in zone 2, 3 and storage tank is faulty and characterizes the sensor $S^{(1)}$ possibly faulty. $I^s$ and $I^{(4)}$ are equal to 0 for all $t > 20.33 h$ and $t > 20 h$, excluding the propagation of sensor faults. The agents $M^{(2)}$ and $M^{(3)}$ do not detect any sensor faults, since the effects of sensor faults occurred in the storage tank and zone 4 on the analytical redundancy relations of $M^{(2)}$ and $M^{(3)}$ where not high enough to be detectable.

5. CONCLUDING REMARKS

In this work, a bank of local sensor fault diagnosis (LSFD) agents have been designed to detect and isolate single and multiple sensor faults. Each LSFD agent consists a local sensor fault detection scheme and a local sensor fault isolation scheme in order to obtain detection and isolation signals, respectively, based on analytical redundancy relations. The detection and isolation signals are retrieved to a local decision logic to determine the location of the sensor faults. Based on the proposed distributed sensor fault diagnosis scheme, future work will focus on the design of a global decision logic for isolating propagated sensors faults based on the (strong) detectability of propagated sensor faults, and a distributed fault tolerant scheme for accommodating the sensor fault effects. An important open issue in fault diagnosis for HVAC systems is the design of diagnostic scheme that is capable of distinguishing between sensor and process faults (e.g., fouled heat exchangers, leaking valves) in HVAC systems.

REFERENCES

ASHRAE (1996). HVAC systems and equipment. American Society of Heating, Refrigerating, and Air Conditioning Engineers, Atlanta, GA.

Bahel, V., Caillat, J.L.M., Herroon, G.P., Hickey, M., Millet, H., Pham, H., and Warner, W.R. (1995). Heat pump motor optimization and sensor fault detection. US Patent 5,440,895.

Balaras, C.A., Dascalaki, E., and Gaglia, A. (2007). HVAC and indoor thermal conditions in hospital operating rooms. Energy and Buildings, 39(4), 454–470.

Dexter, A.L. and Ngo, D. (2001). Fault diagnosis in air-conditioning systems: a multi-step fuzzy model-based approach. HVAC&R Research, 7(1), 83–102.

Du, Z. and Jin, X. (2007). Detection and diagnosis for sensor fault in HVAC systems. Energy Conversion and Management, 48(3), 693–702.

Farrell, J.A. and Polycarpou, M.M. (2006). Adaptive approximation based control: Unifying neural, fuzzy and traditional adaptive approximation approaches, volume 48. Wiley-Interscience.
Gertler, J. (1998). Fault Detection and Diagnosis in Engineering Systems. CRC Press.

Hagras, H., Callaghan, V., Colley, M., and Clarke, G. (2003). A hierarchical fuzzy–genetic multi-agent architecture for intelligent buildings online learning, adaptation and control. Information Sciences, 150(1), 33–57.

Hou, Z., Lian, Z., Yao, Y., and Yuan, X. (2006). Data mining based sensor fault diagnosis and validation for building air conditioning system. Energy Conversion and Management, 47(15), 2479–2490.

Isermann, R. (2006). Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance. Springer Verlag.

Kumar, M. and Kar, I. (2013). Fault detection and diagnosis of air-conditioning systems using residuals. Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems.

Lee, W.Y., House, J.M., and Kyong, N.H. (2004). Subsystem level fault diagnosis of a building’s air-handling unit using general regression neural networks. Applied Energy, 77(2), 153–170.

Liang, J. and Du, R. (2007). Model-based fault detection and diagnosis of HVAC systems using support vector machine method. International Journal of Refrigeration, 30(6), 1104–1114.

Pérez-Lombard, L., Ortiz, J., and Pout, C. (2008). A review on buildings energy consumption information. Energy and Buildings, 40(3), 394–398.

Reppa, V., Papadopoulos, P., Polycarpou, M.M., and Panayiotou, C.G. (2013). Distributed Detection and Isolation of Sensor Faults in HVAC Systems. In Proceedings of Mediterranean Conference on Control and Automation (MED), 401–406.

Reppa, V., Papadopoulos, P., Polycarpou, M.M., and Panayiotou, C.G. (2014). A distributed architecture for HVAC sensor fault detection and isolation. IEEE Transactions on Control Systems Technology, (doi: 10.1109/TCST.2014.2363629).

Schein, J., Bushby, S.T., Castro, N.S., and House, J.M. (2006). A rule-based fault detection method for air handling units. Energy and Buildings, 38(12), 1485–1492.

Wang, S. and Xiao, F. (2004). AHU sensor fault diagnosis using principal component analysis method. Energy and Buildings, 36(2), 147–160.

Wang, S., Zhou, Q., and Xiao, F. (2010). A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults. Energy and Buildings, 42(4), 477–490.

Yu, B., Van Paassen, D.H., and Riahy, S. (2002). General modeling for model-based FDD on building HVAC system. Simulation Practice and Theory, 9(6), 387–397.

Zaheer-Uddin, M. (1994). Temperature control of multi-zone indoor spaces based on forecast and actual loads. Building and Environment, 9(6), 387–397.

Zaheer-Uddin, M. and Patel, R. (1995). Optimal tracking control of multi-zone indoor environmental spaces. Journal of Dynamic Systems, Measurement, and Control, 117(3), 292–303.