

# Decentralized Fault Diagnosis and Prognosis Scheme for Interconnected Nonlinear Discrete-Time Systems

H. Ferdowsi<sup>1</sup> and S. Jagannathan<sup>2</sup>

<sup>1</sup>*Texas A&M University – Texarkana, Texarkana, TX, 75503, USA  
hasan.ferdowsi@tamut.edu or hasan.ferdowsi@mst.edu*

<sup>2</sup>*Missouri University of Science and Technology, Rolla, MO, 65409, USA  
sarangap@mst.edu*

## ABSTRACT

This paper deals with the design of a decentralized fault diagnosis and prognosis scheme for interconnected nonlinear discrete-time systems which are modelled as the interconnection of several subsystems. For each subsystem, a *local fault detector* (LFD) is designed based on the dynamic model of the local subsystem and the local states. Each LFD consists of an observer with an online neural network (NN)-based approximator. The online NN approximators only use local measurements as their inputs, and are always turned on and continuously learn the interconnection as well as possible fault function. A fault is detected by comparing the output of each online NN approximator with a predefined threshold instead of using the residual. Derivation of robust detection thresholds and fault detectability conditions are also included. Due to interconnected nature of the overall system, the effect of faults propagate to other subsystems, thus a fault might be detected in more than one subsystem. Upon detection, faults local to the subsystem and from other subsystems are isolated by using a central fault isolation unit which receives detection time information from all LFDs. The proposed scheme also provides the time-to-failure or remaining useful life information by using local measurements. Simulation results provide the effectiveness of the proposed decentralized fault detection scheme.

## 1. INTRODUCTION

Several practical systems such as the well-known power generation and distribution systems, telecommunication networks, water distribution networks, traffic networks, exhibit complex and spatially distributed dynamics and can be referred to as large-scale interconnected systems. With increasing complexity with these systems, there is a high possibility of occurrence of faults. Therefore, suitable fault diagnosis schemes that help the reliable operation of such

interconnected systems at all times are needed. In this paper, a quantitative decentralized fault diagnosis scheme for a large-scale interconnected system in discrete-time is introduced and its rigorous analysis is included.

Out of the data-driven and model-based fault diagnosis framework, data driven methods (Dash & Venkatasubramanian, 2000) need healthy and faulty data from the system, which can be quite expensive to collect, store and process. Model-based fault diagnosis schemes (Isermann, 2005; Sampath, Sengupta, Lafortune, Sinnamohideen, & Teneketzis, 1995) and model based prognostics (Daigle & Goebel, 2009; Kulkarni, Daigle, Gorospe, & Goebel, 2014; Luo, Namburu, Pattipati, Qiao, Kawamoto, & Chigusa, 2003), on the other hand, do not require significant quantities of data for development and only require data to detect faults online. Therefore, a number of researchers have worked on model-based FD schemes, using adaptive estimators or observers (Demetriou & Polycarpou, 1998; Ferdowsi & Jagannathan, 2013; Wang & Daley, 1996), neural network (NN) based observers (Bernieri, D'Apuzzo, Sansone, & Savastano, 1994; Maki & Loparo, 1997), fuzzy observers (Blake & Brown, 2007; Patton, Chen, & Lopez-Toribio, 1998) and so on, for several practical industrial systems. However, these schemes are centralized and not suitable for large scale systems with multiple distributed subsystems. In such systems, measurements are taken at a subsystem need to be transmitted to all subsystems, which is not appropriate since continuous transmission of large amounts of data over the entire distributed system is both costly and prone to errors and delays. This is the main motivation behind the development of decentralized methods for control and diagnostics.

While traditional fault diagnosis articles (Bernieri et al., 1994; Blake & Brown, 2007; Demetriou & Polycarpou, 1998; Ferdowsi & Jagannathan, 2013; Maki & Loparo, 1997) offer only centralized FD schemes that require the entire state vector of the system to be measured and transmitted, in the recent literature, decentralized control of distributed systems (Boskovic & Mehra, 2002; Huang, Tan, & Lee, 2005, 2006;

---

Hasan Ferdowsi is the corresponding author. Research supported in part by NSF I/UCRC on Intelligent Maintenance Systems and Intelligent Systems Center at Missouri University of Science and Technology. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Ferdowsi & Jagannathan 2017) by using local subsystem state vector is introduced. By using overlapping decomposition (Ferrari, Parisini, & Polycarpou, 2009), a large-scale system is decomposed into a set of subsystems which are connected by unknown nonlinear functions. Then a distributed fault diagnosis scheme is introduced by assuming that the entire state vector is available. On the other hand, decentralized fault diagnosis schemes in (Ferrari et al., 2009; Stankovic, Ilic, Djurovic, Stankovic, & Johansson, 2010) are introduced for continuous-time systems by assuming that the interconnection functions are known and the entire estimated system state vector is available at each subsystem. However, for large-scale interconnected systems, it is very expensive and time consuming to transmit measured or estimated states to all subsystems to aid in an accurate diagnosis and even if the required hardware is already in place, such transmissions can be delayed and prone to errors.

On the contrary, our objective in this paper is to design a network of local fault detectors (LFD) or observers for interconnected nonlinear discrete-time systems so that each LFD monitors a single subsystem by making use of the local information or state vector in contrast with (Ferrari et al., 2009; Stankovic et al., 2010). In addition, partial isolation of faults and TTF estimation will be performed upon detection, which provide further advantages over the existing schemes (Bernieri et al., 1994; Blake & Brown, 2007; Ferrari et al., 2009; Stankovic et al., 2010).

Since discrete-time implementation is preferred for hardware implementation (Caccavale & Villani, 2004), in this work, a nonlinear discrete-time system is considered with external disturbances, unmodeled dynamics, interconnection effects, and nonlinear fault functions that cover both abrupt and incipient faults. Incipient faults may be difficult to deal with due to the fact that their small effects on residuals can be hidden as if they are due to the modeling uncertainty. Here, we stress the design of truly decentralized fault diagnosis scheme in discrete-time for incipient faults.

As mentioned above, one local fault detector is designed for each subsystem. Each LFD mainly consists of a nonlinear observer with an online NN approximator which is used to estimate the unknown part of the subsystem dynamics, i.e. interconnection and possible fault functions, by using only the local state information. It is mathematically shown that although the interconnection term is a function of nonlocal state vector, it can be estimated by an online approximator whose inputs are the measured local states at the current and previous time instant. The history of local state vector will help overcome the need for the entire system state vector.

A local residual signal is generated by comparing the estimated local state vector from the observer with the measured subsystem state vector. However, this residual is not used for performing fault detection, whereas it is used to update the unknown parameters of the online NN approximator. In contrast with other model-based fault detection methods (Bernieri et al., 1994; Blake & Brown, 2007; Demetriou & Polycarpou, 1998; Ferdowsi &

Jagannathan, 2013; Ferrari et al., 2009; Maki & Loparo, 1997; Patton et al., 1998; Stankovic et al., 2010; Wang & Daley, 1996), the online approximator is always active and the detection is performed by comparing the output of the online NN approximator in discrete-time (OLAD) with a predefined threshold. This is possible due to the fact that the interconnection term remains bounded as long as the system is healthy with no fault present since the system state vector remain bounded in the absence of fault due to the presence of a stabilizing controller. In addition, a mathematically rigorous approach to the derivation of robust detection thresholds and fault detectability condition is given.

The approximation of interconnection and fault functions allows a good estimation of state vector, thus allowing proper estimation of TTF by comparing the system state estimate vector against the user defined failure limits (Thumati & Jagannathan, 2010). The TTF can help ensuring that the system will not be operated beyond this limit as it is unsafe. In this paper, the TTF is determined by using estimated system state vector instead of parameter estimate vector.

Upon detection, a fault isolation algorithm is utilized to determine whether or not the fault is local by making use of a central fault isolation unit. Under the assumption that the local faults affect local measurements quicker than non-local faults, the location of the fault is identified by comparing the detection times from all LFDs. Note that the detection and prediction units are purely decentralized since they are independent of the isolation unit.

In our earlier and preliminary work (Ferdowsi, Raja, & Jagannathan, 2012a; Ferdowsi, Raja, & Jagannathan, 2012b), the fault diagnosis is addressed for interconnected systems where it is assumed that the interconnection terms are bounded as a function of state and parameter estimation errors which is a stringent assumption. In contrast, in the current work the interconnection terms are estimated online which allows determination of fault effects on other subsystems. Consequently, the fault detection is performed differently by using the OLAD outputs rather than residual. Moreover, the TTF estimation in (Ferdowsi et al., 2012a) was directly borrowed from (Thumati & Jagannathan, 2010) by using the magnitude of the fault parameters, whereas the TTF estimation in this paper is based on the observer dynamics instead of weight update law since failure limits cannot be determined for the NN weights. Rigorous convergence analysis, analytically derived detection threshold, and detectability condition are other contributions of this paper over (Ferdowsi et al., 2012a).

Thus the major contributions of this paper include the development of a decentralized fault diagnosis scheme for nonlinear discrete-time systems wherein a LFD only uses local measurements in contrast with (Bernieri et al., 1994; Blake & Brown, 2007; Demetriou & Polycarpou, 1998; Ferdowsi & Jagannathan, 2013; Maki & Loparo, 1997; Patton et al., 1998; Wang & Daley, 1996). Here the interconnection term is not neglected in contrast with (Ferdowsi et al., 2012a). Furthermore, the TTF estimation is performed upon fault

detection by using the observer dynamics whereas such scheme is not available in existing methods (Ferrari et al., 2009; Stankovic et al., 2010) for interconnected systems.

This paper is organized as follows. Section 2 gives a system description for interconnected nonlinear discrete-time systems. Section 3 proposes the decentralized fault detection scheme, and discusses the partial isolation of faults as well as TTF determination, and Section 4 reports simulation results. An automated highway system is used as a running example throughout the paper to better illustrate the design steps.

## 2. SYSTEM DESCRIPTION

Consider the interconnected nonlinear discrete-time systems described by

$$x(k+1) = F(x(k), u(k)) + \eta(x(k), u(k)) + h(x(k), u(k)),$$

where  $u \in \mathbb{R}^m$  is the control input vector,  $x \in \mathbb{R}^n$  is the system state vector,  $F: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$  represents the nonlinear system dynamics,  $\eta: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$  represents the system uncertainties, and  $h: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$  represents a vector of possible fault dynamics. Suppose that this system is comprised of  $N$  interconnected subsystems. The  $i^{\text{th}}$  subsystem dynamics are given by

$$x_i(k+1) = f_i(x_i(k), u_i(k)) + g_i(x_i(k), \bar{x}_i(k), u_i(k)) + \eta_i(x_i(k), u_i(k)) + h_i(x_i(k), u_i(k)), \quad (1)$$

where  $u_i \in \mathbb{R}^{m_i}$  is the local control input vector,  $x_i \in \mathbb{R}^{n_i}$  is the local state vector,  $\bar{x}_i \in \mathbb{R}^{\bar{n}_i}$  is the non-local state vector ( $\bar{x}_i$  includes all system states except for those in  $x_i$ , which means  $x_i$  and  $\bar{x}_i$  are mutually exclusive and  $x = x_i \cup \bar{x}_i$ ),  $f_i: \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$  and  $g_i: \mathbb{R}^{n_i} \times \mathbb{R}^{\bar{n}_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$  represent the known local and unknown interconnection functions respectively,  $\eta_i: \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$  denotes the system uncertainties, and  $h_i: \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{n_i}$  is the local fault function or fault dynamics.

The fault function  $h_i$  can obviously represent an abrupt fault. However, in order to cover a wider range of faults,  $h_i$  can be expressed as the multiplication of a time profile and the fault magnitude, i.e.  $h_i = \Pi_i(k - k_0) \bar{h}_i$ . The time profile  $\Pi_i(k - k_0)$  is modeled by

$$\Pi_i(k - k_0) = \text{diag} \{ \Omega_{i_1}(k - k_0), \Omega_{i_2}(k - k_0), \dots, \Omega_{i_{n_i}}(k - k_0) \},$$

where  $\Omega_{i_j}(\tau) = \begin{cases} 0, & \text{if } \tau < 0 \\ 1 - e^{-\bar{\kappa}_j \tau}, & \text{if } \tau \geq 0 \end{cases}$  for  $j = 1, \dots, n_i$ , is the

time profile and  $\bar{\kappa}_j$  is an unknown constant that represents the rate at which a fault develops. A larger value of  $\bar{\kappa}_j$  indicates that the fault has a larger growth rate. The use of such time profiles is common in fault diagnosis literature (Thumati & Jagannathan, 2010; Zhang & Morris, 1994). Next the standard assumptions are needed in order to proceed.

Assumption 1 (Demetriou & Polycarpou, 1998): The modeling uncertainty is locally bounded, i.e.  $\|\eta_i(x_i(k), u_i(k))\| \leq \eta_{i_M}, \forall (x_i, u_i) \in S_i, i = 1, 2, \dots, N$ ,

where  $\eta_{i_M}$  is a positive known constant and  $S_i$  is a region to be defined in assumption 4.

Remark 1: Assumption 1 is needed to distinguish between faults and system uncertainties.

Assumption 2 (Thumati & Jagannathan, 2010): Both interconnection and fault functions are expressed as nonlinear in the unknown parameters (NLIP) (Jagannathan, 2006), i.e. they can be approximated by two-layer NN with bounded activation functions and weight parameters.

Assumption 3 (Huang et al., 2005): The interconnection terms are bounded by polynomial-type nonlinearities as  $\|g_i(x(k), u_i(k))\| \leq \sum_{j=1}^N (\zeta_{ij}^0 + \zeta_{ij}(x_j))$ ,  $i = 1, 2, \dots, N$ , where  $\zeta_{ij}^0$  is constant and  $\zeta_{ij}(\cdot)$  is a bounded smooth function for  $i, j = 1, 2, \dots, N$ .

Assumption 4 (Demetriou & Polycarpou, 1998): There exists a stabilizing controller that guarantees the boundedness of system state vector during the healthy conditions. This implies that  $(x_i, u_i) \in S_i$  where  $S_i$  is a bounded region.

Remark 2: By combining Assumptions 3 and 4, it can be shown that, under healthy operating conditions, the interconnection term is bounded by  $\|g_i(x(k), u_i(k))\| \leq g_{i_M}$ ,  $i = 1, 2, \dots, N$ , where  $g_{i_M}$  is a positive constant. This result is only used to identify the fault detection thresholds as it is not valid under faulty condition. On the other hand, during fault conditions, the OLAD to be defined in the next section approximates the interconnection function as it propagates the fault from one subsystem to another.

Assumption 5: All system states are measurable.

Assumption 6: Only a single fault can exist in the system at any given time.

Remark 3: Assumption 6 is only required for the isolation part and it is not needed for fault detection or failure prediction.

As mentioned in the introduction, a running example will be provided throughout the paper for improved clarity. An automated highway system (Yan & Edwards, 2008), is selected for this purpose. Each vehicle will be considered as one subsystem. The dynamics for the  $i^{\text{th}}$  vehicle is given by

$$\begin{cases} \psi_i(k+1) = \psi_i(k) + T(v_i(k) - v_{i-1}(k)) \\ v_i(k+1) = v_i(k) + \frac{T}{M_i}(-D_i v_i^2(k) + \xi_i(k) - d_i) \\ \xi_i(k+1) = \xi_i(k) + \frac{T}{\tau_i}(\delta_i(k) - \xi_i(k)) \end{cases} \quad (2)$$

where  $\delta_i$  is the control input for  $i^{\text{th}}$  vehicle,  $\xi_i$  is the driving/braking force applied to the  $i^{\text{th}}$  vehicle,  $v_i$  is the velocity of the  $i^{\text{th}}$  vehicle,  $\psi_i$  is the distance between vehicle  $i$  and the vehicle in front of it (note that based on its definition,  $\psi$  will be negative),  $T$  is the sampling time, and  $M_i$ ,  $D_i$ ,  $d_i$ , and  $\tau_i$  are the  $i^{\text{th}}$  vehicle's mass, aerodynamic drag, frictional force, and engine/brake time constant respectively. Moreover, a virtual leader is defined for the first vehicle to follow. The control objective is to make each vehicle follow the vehicle in front of it with a safe distance. This problem is formulated to regulation of  $y_i = \psi_i + L + P v_i$  to zero (Yan & Edwards, 2008), where  $L$  and  $P$  are positive constants that determine the

required distance at zero speed and velocity-dependent distance ratio. Typical values for  $L$  and  $P$  are 1 and 0.9 respectively. A backstepping controller designed to satisfy the requirements is as follows:

$$d_i(k) = u_i(k) + \bar{u}_i(k)$$

where

$$u_i(k) = \tau_i \left[ \frac{\xi_i(k)}{\tau_i} - \frac{M_i}{P} (2v_i(k) + \psi_i(k) + L + Pv_i(k)) + \left( \frac{2D_i}{M_i} v_i(k) - \frac{1}{P} - 2 \right) (-D_i v_i^2(k) + \xi_i(k) - d_i) - \frac{d_{i-1}}{P} \right]$$

$$\bar{u}_i(k) = \tau_i \left[ \frac{2M_i}{P} v_{i-1}(k) + \frac{1}{P} (-D_{i-1} v_{i-1}^2(k) + \xi_{i-1}(k)) \right]$$

Note that  $u_i$  is only a function of states of vehicle  $i$ , while  $\bar{u}_i$  includes states of the vehicle ahead. The system dynamics in (2) is in the form of system description in (1) if the local and interconnection terms are defined as

$$f_i(x_i, u_i) = \begin{bmatrix} \psi_i(k) + Tv_i(k) \\ v_i(k) + \frac{T}{M_i} (-D_i v_i^2(k) + \xi_i(k) - d_i) \\ \xi_i(k) + \frac{T}{\tau_i} (u_i(k) - \xi_i(k)) \\ -Tv_{i-1}(k) \\ 0 \\ T \left( \frac{2M_i}{P} v_{i-1}(k) + \frac{1}{P} (-D_{i-1} v_{i-1}^2(k) + \xi_{i-1}(k)) \right) \end{bmatrix} \quad (3)$$

where  $x_i = [\psi_i \ v_i \ \xi_i]^T$  is the state vector of subsystem  $i$ . Several different faults can be investigated for the automated highway system. For example partial loss of actuation in the  $i^{\text{th}}$  vehicle which can be induced in simulations by the following fault function

$$h_i = \begin{bmatrix} 0 \\ \frac{T}{m_i} (-b\xi_i(k)) \\ 0 \end{bmatrix}$$

where  $0 < b < 1$  is the ratio of loss in actuation. Modeling uncertainty and noise will also be added in the simulations (refer to section 4) to make the problem even more realistic.

### 3. FAULT DIAGNOSIS AND PROGNOSIS SCHEME

Next the proposed fault detection scheme is introduced.

#### 3.1. Fault Detection (FD)

In order to monitor the system state vector, estimators using local measurements are designed. Since the interconnection and fault functions are not known, an NN-based online approximator in discrete time, referred to as OLAD, is incorporated in each local estimator to approximate these functions. Unlike other fault detection schemes where the OLAD is turned on only after the detection, the OLADs used in our proposed estimators are always turned on, in order to learn the possible fault dynamics as well as the interconnection dynamics.

Let  $\omega_i$  be defined as the summation of interconnection term and fault function in subsystem  $i$  as

$$\omega_i(k) = g_i(x_i(k), \bar{x}_i(k), u_i(k)) + \Pi(k - k_0) h_i(x_i(k), u_i(k)).$$

It is clear, based on (1), that the interconnection term at time  $k$ , will affect the local state vector at the next time instant  $k+1$ . Using this fact, the interconnection term at time  $k$  can be represented as a function of local state vector at time  $k+1$  and local state and input vector at time  $k$ . Thus,  $\omega_i(k)$  can be approximated by an online approximator such as a two layer neural network (NN) whose inputs consist of  $x_i(k)$ ,  $u_i(k)$ , and  $x_i(k+1)$ , with bounded weights and approximation error, i.e.  $\omega_i(k) = \theta_i^T(k) \phi_i(x_i(k+1), x_i(k), u_i(k)) + \varepsilon_i(k)$ , where  $\theta_i(k)$  is the unknown parameter matrix,  $\phi_i(x(k), u(k))$  is a basis function like sigmoid, and  $\varepsilon_i(k)$  is the approximation error which is bounded by  $\varepsilon_{iM}$ . However, since the measured state vector,  $x_i(k+1)$ , is not available at time  $k$ , we will consider the online approximator one time step behind the actual system, in order to make the proposed scheme practical. Thus, the OLAD will be incorporated in the nonlinear observer which is designed to work one time step behind the actual system. The residual, which is defined as the error between measured and estimated states, will then be used to update the NN weights.

Consider the local nonlinear estimator for the  $i^{\text{th}}$  subsystem described by

$$\hat{x}_i(k) = \lambda \hat{x}_i(k-1) + f_i(x_i(k-1), u_i(k-1)) - \lambda x_i(k-1) + \hat{\omega}_i(x_i(k), x_i(k-1), u_i(k-1); \hat{\theta}_i(k-1)), \quad (4)$$

for  $k \geq 1$ , where  $\hat{x}_i(k) \in \mathbb{R}^{n_i}$  is the estimated local state vector of the  $i^{\text{th}}$  subsystem,  $\hat{\omega}_i: \mathbb{R}^{m_i} \times \mathbb{R}^{n_i} \times \mathbb{R}^{p_i \times n_i} \rightarrow \mathbb{R}^{n_i}$  is the output of the OLAD with  $\hat{\theta}_i \in \mathbb{R}^{p_i \times n_i}$  being its set of unknown parameters and  $\lambda$  is a user defined constant, which must be selected in a way that the eigenvalues of the closed loop system lie within the unit disc (Jagannathan, 2006). Initial values of the local fault detection (FD) estimator are taken as  $\hat{x}_i(0) = \hat{x}_{i0}$ ,  $\hat{\theta}_i(0) = \hat{\theta}_{i0}$ .

During the healthy operating condition of the system, the following holds  $\|\omega_i(k)\| \leq \|g_i(x(k), u_i(k))\|$  so that  $\|\omega_i(k)\|$  remains bounded based on Assumptions 3. When a fault occurs in a subsystem, the magnitude of the fault function in the subsystem and the magnitude of the interconnection term in the other subsystems will increase. Therefore, a fault can be detected by comparing the norm of OLAD output,  $\|\hat{\omega}_i\|$ , with a detection threshold  $\rho_i$  which will be defined later by using the bound on the interconnection functions in the healthy operating conditions as well as the bound on the OLAD approximation error. This is in contrast with detecting a fault by using the residual or state estimation error. In this method, the residual cannot be used for fault detection since it will always remain close to zero because the OLAD is always online and it will estimate the unknown part of the subsystem dynamics.

To move forward, define the  $i^{\text{th}}$  subsystem residual as  $e_i(k) = x_i(k) - \hat{x}_i(k)$ . Prior to the occurrence of a fault, the local residual dynamics are obtained by comparing (1) and (4), as

$$e_i(k) = \lambda e_i(k-1) + g_i(x(k-1), u_i(k-1)) + \eta_i(x_i(k-1), u_i(k-1)) - \hat{\omega}_i(x_i(k), x_i(k-1), u_i(k-1); \hat{\theta}_i(k-1)), \quad (5)$$

where  $x$  is the entire state vector, i.e.  $x = \bigcup_{i=1}^N x_i$ . The next step in the design is to determine an update law for the OLAD. Define the parameter update law of the OLAD as

$$\hat{\theta}_i(k) = \hat{\theta}_i(k-1) + \alpha_i \phi_i(k-1) e_i^T(k-1) - \gamma_i \|I - \alpha_i \phi_i(k-1) \phi_i^T(k-1)\| \hat{\theta}_i(k-1), \quad (6)$$

where  $\alpha_i > 0$  is the learning rate,  $0 < \gamma_i < 1$  is the forgetting factor, and  $\phi_i(k) = \phi_i(x_i(k), u_i(k))$  is a basis function such as sigmoid or radial basis functions (RBF). Then, the output of the OLAD is calculated as

$$\hat{\omega}_i(k-1) = \hat{\theta}_i^T(k-1) \phi_i(x_i(k), x_i(k-1), u_i(k-1)). \quad (7)$$

Upon detection the local error dynamics can be derived by comparing (1) and (4) at time  $k$  as

$$e_i(k) = \lambda e_i(k-1) + g_i(x(k-1), u_i(k-1)) + \eta_i(x_i(k-1), u_i(k-1)) + h_i(x_i(k-1), u_i(k-1)) - \hat{\omega}_i(x_i(k), x_i(k-1), u_i(k-1); \hat{\theta}_i(k-1)). \quad (8)$$

Asserting the NLIP assumption on the local fault function, the above equation can be rewritten as

$$e_i(k) = \lambda e_i(k-1) + \eta_i(x_i(k-1), u_i(k-1)) + \tilde{\theta}_i^T(k-1) \phi_i(x_i(k), x_i(k-1), u_i(k-1)) + \varepsilon_i(k-1), \quad (9)$$

where  $\tilde{\theta}_i(k) = \theta_i(k) - \hat{\theta}_i(k)$  represents the parameter estimation error and  $\varepsilon_i(k)$  is the OLAD approximation error, which is bounded by  $\varepsilon_{iM}$  due to Assumption 2. Next the stability of the local FD residual and parameter estimation errors is discussed.

**Theorem 1 (Local Fault Detection Observer Performance):** Let the proposed local FD observer defined in (4) be used to monitor the subsystem described by (1), and let the update law in (6) be used to update the unknown parameter vector,  $\hat{\theta}_i(k)$ . In the presence of system uncertainties and under the Assumptions 1 through 4, the local FD residual,  $e_i(k)$ , and the parameter estimation error,  $\tilde{\theta}_i(k)$ , are uniformly ultimately bounded, provided the user-defined constants,  $\lambda$  and  $\alpha_i$ , and  $\gamma_i$ , are selected such that  $|\lambda| < 0.5$ ,  $\alpha_i < \sqrt{(1 - 4\lambda^2)/24\phi_{i_{max}}^4}$ , and  $0.6 < \gamma_i < \frac{1.4}{1 + \alpha_i \phi_{i_{max}}^2}$ .

**Proof:** Refer to the appendix.

Theorem 1 guarantees the stability of the local FD residual and parameter estimation errors provided the design parameters are selected as per Theorem 1. When a fault happens in a subsystem, the output of the OLAD in that subsystem will include an approximation of the fault function in addition to the interconnection term, while the OLADs in other subsystems will only approximate interconnection functions. Although the fault function only exists in one of the

subsystems, it will affect the other subsystems through the interconnection terms. Therefore, the estimation of interconnection functions in non-faulty subsystems, allows determination of non-local fault effects.

Based on Assumptions 3 and 4, the interconnection terms are bounded during healthy conditions and based on Theorem 1, the OLADs approximate the interconnection terms with bounded error during healthy conditions. Therefore, OLAD outputs are bounded as long as the system is working under healthy conditions. This result is used in the next theorem to show that no false alarms will be generated if the detection thresholds are selected appropriately. Consequently, a detectability condition is analytically derived that guarantees the detection of faults which can satisfy this condition.

**Theorem 2 (Robustness and Detectability):** Consider the nonlinear subsystem defined by (1) and the local observer (4). No fault is detected under healthy operating conditions if the detection threshold is selected as

$$\rho_i = g_{iM} + \varepsilon_{iM} + \phi_{i_{max}} \sqrt{D/C_2} + q_i, \quad (10)$$

where  $q_i$  is a user-defined small positive constant, and  $D$  and  $C_2$  are defined in the appendix. On the other hand, the fault in subsystem  $i$  will be detected by its local fault detector, if there exists a time instant  $k_d$ , at which the following condition on the fault function is satisfied

$$\|h_i(x_i(k_d), u_i(k_d))\| \geq 2\rho_i - q_i. \quad (11)$$

**Proof:** Refer to the appendix.

To better illustrate the fault detection method, the running example is revisited here. A local estimator should be developed for each vehicle in the automated highway system introduced in section 2. The observer for the  $i^{\text{th}}$  vehicle can be obtained based on (4) as follows

$$\hat{x}_i(k) = \lambda \hat{x}_i(k-1) + f_i(x_i(k-1), u_i(k-1)) - \lambda x_i(k-1) + \hat{\omega}_i(k-1), \quad (12)$$

Recall that  $x_i = [\psi_i \ v_i \ \xi_i]^T$  is the state vector of vehicle  $i$  and  $f_i(\cdot)$  is provided in (3).  $\hat{\omega}_i$  is the online approximator created by a two layer neural network with 10 neurons in the hidden layer. For this example, a sigmoid function is used as basis function, i.e.  $\phi_i(z) = 1/(1 + e^{-z})$ , thus  $\hat{\omega}_i$  is calculated by

$$\hat{\omega}_i(k-1) = \hat{\theta}_i^T(k-1) \phi_i(A_N [x_i(k) \ x_i(k-1) \ u_i(k-1)]^T + B_N) \quad (13)$$

where the matrices  $A_N \in \mathbb{R}^{10 \times 7}$  and  $B_N \in \mathbb{R}^{10 \times 1}$  are selected randomly and  $\hat{\theta}_i \in \mathbb{R}^{10 \times 3}$  is the weight matrix initiated at zero. The weights start updating by equation (6) as soon as the observer starts working.

To select the observer and update law parameters based on theorem 1,  $\phi_{i_{max}}$  is needed. The maximum value of the selected basis function  $\phi_i$  is one and  $\phi_i$  is a  $10 \times 1$  vector, therefore  $\phi_{i_{max}} = \sqrt{10}$ . Therefore, the observer and update law parameters should be selected such that

$$|\lambda| < 0.5$$

$$\alpha_i < \sqrt{(1 - 4\lambda^2)/2400}$$

$$0.6 < \gamma_i < \frac{1.4}{1 + 10\alpha_i}$$

For this example the parameters were selected as  $\lambda = 0.01$ ,  $\alpha_i = 0.01$ , and  $\gamma_i = 0.7$  to satisfy the above conditions.

Next section will discuss the partial fault isolation, upon detection of a fault by a local FD. With the proposed isolation method, the detected fault can be characterized as local or non-local fault to each subsystem.

### 3.2. Fault Isolation

As discussed earlier, the OLAD in faulty subsystem estimates both the local fault function and the interconnection term, while the OLADs in other subsystems estimate only their interconnection function, which could be affected by the nonlocal fault. Therefore, not only the output of OLAD where the fault has occurred will increase above the detection threshold, but also the outputs of other OLADs can possibly increase due to interconnection effects. Thus, detection of a fault might happen in more than one subsystem.

In this context, local and non-local faults should be defined. A fault that has occurred in subsystem  $i$ , will be called a local fault for this subsystem and it will be referred to as a non-local fault for other subsystems. Under the assumption that local faults affect local measurements quicker than the non-local faults due to smaller propagation delay, a heuristic fault isolation algorithm is developed based on the detection times in all subsystems. In the proposed isolation method, communication between the LFD and the centralized isolation unit is required. However, there is no need for the transmission of the measured or estimated state vector of all the subsystems at each sampling interval. Also, note that the detection information must only be transmitted when a fault is detected, which means no transmission is needed in healthy operating condition which is the majority of the time. The only information that must be transmitted after detection is the detection time in each local fault detector and there is no need for the detection information to be transmitted at each and every time instant. In fact this information must be sent from all the subsystems to a central isolation unit at time instants  $k = jn$  where  $j = 1, 2, \dots$  and  $n$  is a positive integer which determines the rate at which detection information must be collected from all the subsystems.

In other words, the time interval between two consecutive transmissions will be equal to  $nT$  where  $T$  is the sampling time. Larger value of  $n$  will result in fewer number of transmissions over the network, while smaller value of  $n$  leads to faster isolation of faults. So there is a tradeoff here which means that  $n$  should be selected according to both the required isolation speed and preferred transmission interval in a specific system.

To formulate the isolation scheme, let  $t_D^{(i)}$  be the variable used to store the detection information of subsystem  $i$  and let  $t_D^{(i)}(0) = 0$ . The value of  $t_D^{(i)}$  will remain at zero unless a fault is detected by the LFD of subsystem  $i$ . Once a fault is detected

by this LFD,  $t_D^{(i)}$  will be set to the detection time, i.e.,  $t_D^{(i)}(k) = Tk_{detection}^{(i)}$  for  $k \geq k_{detection}^{(i)}$  where  $k_{detection}^{(i)}$  is the time at which a fault is detected by the LFD in subsystem  $i$ . Note that  $t_D^{(i)}(k)$  is sent to the central isolation unit only when  $k = nj$  where  $j = 1, 2, \dots$ .

The fault isolation flowchart is depicted in Figure 1. Once detection information ( $t_D^{(i)}$ ) is sent to the isolation unit by all the subsystems, the minimum among all of the nonzero detection times is calculated. Then for each subsystem, say subsystem  $i$ ,  $t_D^{(i)}$  is first compared to zero. When  $t_D^{(i)}$  is equal to zero obviously no fault has been detected in subsystem  $i$ . However, when  $t_D^{(i)} > 0$  a local or nonlocal fault has been detected in subsystem  $i$ . In this case, if the detection time  $t_D^{(i)}$  is equal to the minimum of all nonzero detection times, then the fault will be isolated local to subsystem  $i$ , otherwise the fault is a nonlocal fault which has propagated to subsystem  $i$ .

Regarding the running example, the isolation algorithm will only be required if a fault is initiated in one vehicle but detected in more than one vehicle. In that case the central isolation unit must receive the detection times and compare them to find out where the fault has been detected first to identify the location of fault.

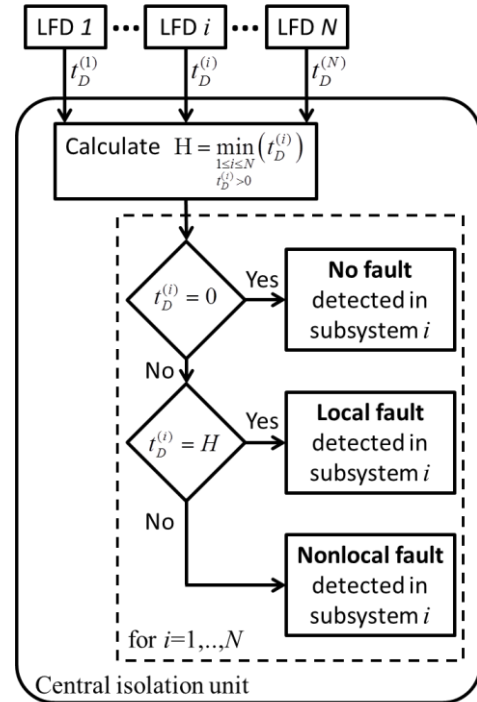


Figure 1: Flowchart of the fault isolation.

Remark 4: Note that with this method of fault detection and isolation, not only the location of fault can be determined, but also all the subsystems which are affected by this fault are identified.

Remark 5: The effectiveness of this isolation method depends on the selection of fault detection thresholds. For

example, if the detection threshold of the subsystem where the fault occurs is significantly larger than the actual bound on the OLAD output of that subsystem in healthy conditions, the fault might be first detected in other subsystems, leading to an incorrect isolation decision. The proposed isolation strategy is most suitable for networks of similar subsystems with same characteristics where one threshold value is appropriate for all subsystems. In that case, all subsystems share the same situation with respect to the detection threshold, which will prevent incorrect isolation decision with the proposed method.

### 3.3. Failure Prediction

The time-to-failure (TTF) determination is necessary for prognostics and this is referred to as remaining useful life of the system. After the detection of a fault, by comparing the estimated state vector obtained from the observer to the user defined limits, time to failure can be determined (Thumati & Jagannathan, 2010). System states represent physical parameters that have failure limits. The TTF is defined as the remaining time until at least one state reaches its limit. As mentioned before, a fault might be detected in more than one subsystem, since any local fault can influence other subsystems as well. Therefore, TTF estimation should be performed for all the subsystems which are significantly affected by the fault, i.e. all subsystems where detection has occurred. The TTF estimation starts in a subsystem immediately after detection.

In order to predict the time of failure, the dynamics of the system can be used which will help determine the rate of change of system states. Since there exist unknown terms in the actual system dynamics (1), the observer dynamics (4) is utilized. According to the stability analysis presented earlier, observer states follow actual states with bounded error which can be decreased by proper selection of design parameters.

Therefore, in the TTF determination, the estimated state dynamics in (4) are utilized to project the estimated state to reach a predefined threshold. The estimated state is driven by the fault approximator. The following theorem provides an analytical formula for finding an estimation of TTF at any time after the detection of a fault. The main idea is introduced in (Thumati & Jagannathan, 2010) for systems that are not distributed in nature. In contrast, the derivation of the formula introduced here is different due to the distributed nature of the system and the fact that the weight update law cannot be used to estimate time-to-failure, because the NN weights do not correspond to real physical parameters and their failure thresholds are not available. This is why the observer dynamics are used instead of the NN weight update law.

**Theorem 3 (TTF Estimation):** Upon detection in subsystem  $i$ , TTF for the  $j^{\text{th}}$  state at the  $k^{\text{th}}$  time instant can be estimated using

$$TTF_{i,j} = \log_{\lambda} \left| \frac{(1-\lambda)x_{i,j_M} - s_{i,j}(k-1)}{\lambda(1-\lambda)\hat{x}_{i,j}(k-1) - \lambda s_{i,j}(k-1)} \right|, \quad (14)$$

where  $x_{i,j_M}$  is the failure threshold of the  $j^{\text{th}}$  state of the  $i^{\text{th}}$  subsystem,  $\hat{x}_{i,j}$  is the estimated value of the corresponding state, and  $s_{i,j}(k-1)$  is the  $j^{\text{th}}$  element of the vector  $s_i(k-1)$  which is defined by

$$s_i(k-1) = f_i(x_i(k-1), u_i(k-1)) - \lambda x_i(k-1) + \hat{\omega}_i(x_i(k), x_i(k-1), u_i(k-1); \hat{\theta}_i(k-1)). \quad (15)$$

**Proof:** Refer to the appendix.

Figure 2 illustrates the process of finding the TTF after a fault is detected in subsystem  $i$ . At each time instant, after calculating the TTF for all the local subsystem parameters, the overall minimum of all TTFs for all of the parameters is calculated to get the overall TTF for the subsystem. This is because the system will be unsafe even if only one of its parameters reaches its limit.

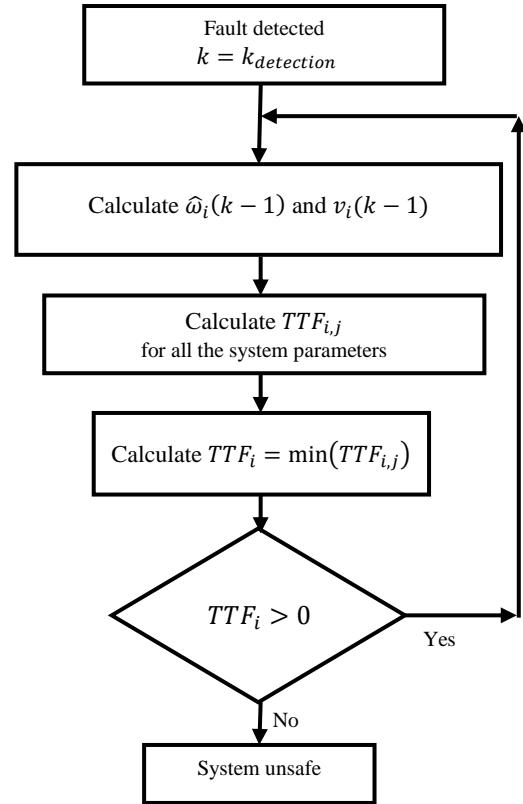


Figure 2: Flow chart of the TTF determination.

The failure prediction scheme can be easily applied to the automated highway system example. Each of the states of vehicle  $i$  have failure limits. For example, the failure limit on each vehicle's distance to the vehicle ahead of it cannot be less than one meter, thus setting a failure limit of -1 on the value of  $\psi_i$ . Similarly, the failure limit on each vehicle's

velocity  $v_i$  and driving force  $\xi_i$  are 30 and 5000 respectively. Since the state vector is defined as  $x_i = [\psi_i \ v_i \ \xi_i]^T$ , the failure limit vector will be  $x_{iM} = [-1 \ 30 \ 5000]^T$ . After the detection of a fault in any vehicle, equation (15) will provide an estimate of the time-to-failure in that subsystem.

#### 4. SIMULATION RESULTS

Two examples are selected to show the effectiveness of the proposed scheme. First, the automated highway system and then a five-tank water system will be presented.

##### 4.1. Automated Highway System

An automated highway system consisting of 50 vehicles is considered here. Each vehicle dynamics is given by (2) and all vehicles start at rest with initial distances of 1 meter. Vehicle parameters are given as  $m_i = 1300\text{kg}$ ,  $D_i = 0.3\text{Ns}^2/\text{m}^2$ ,  $d_i = 100\text{N}$ ,  $\tau_i = 0.2\text{s}$ , and the sampling time is  $T = 0.1\text{s}$ . The desired velocity of the virtual leader is selected as  $v_d(k) = 20 + 0.2 \sin(0.1kT)$  and the backstepping controller presented in section 2 is used to control the vehicles.

The states of the first three vehicles in healthy operating conditions are shown in Figure 3. Next, an abrupt fault in the form of 50% loss of actuation is induced in the second vehicle at time  $t=50\text{s}$  and the second system states are shown in Figure 4. As mentioned in section 2, the fault function is

$$h_2 = \frac{1 + \text{sign}(kT - 50)}{2} \begin{bmatrix} 0 \\ T \\ m_2(-b\xi_2(k)) \\ 0 \end{bmatrix}$$

where  $b = 0.5$ .

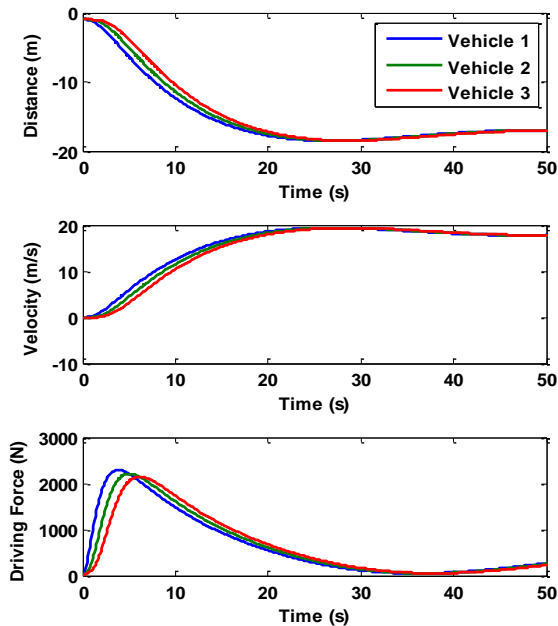


Figure 3: First three vehicle states in healthy conditions

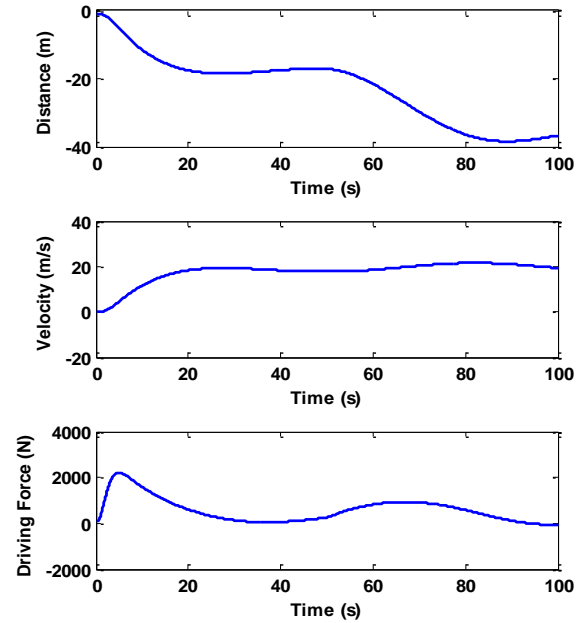


Figure 4: Actual second vehicle states in faulty conditions

To apply the proposed scheme on this system, each vehicle is considered as one subsystem for which the observer in (12) along with the online approximator in (13) are used for fault detection. As mentioned in section 3, the observer and approximator parameters were selected as  $\lambda = 0.01$ ,  $\alpha_i = 0.01$ , and  $\gamma_i = 0.7$  to satisfy the stability conditions. In order to simulate the real world situation, modeling uncertainty and noise are added to the model and measurements used by the observer. Uncertainty is created by a 2% inaccuracy in engine time constant for the model used in equation (12).

The estimated states of the second vehicle are shown in Figure 5 and the state estimation error is shown in Figure 6. It can be observed from Figure 6 that the state estimation error does not increase after the occurrence of fault which is due to the fact that the online approximator is always active and estimating the unknown parts of the system dynamics, namely the interconnections, uncertainty, and fault. Obviously, the state estimation error will not be used for fault detection. As discussed in section 3, fault detection is performed by comparing the norm of the online approximator output in each subsystem with the detection threshold, which is selected by using equation (10). The norm of online approximator outputs for the first three vehicles are given in Figure 7 along with the detection threshold. As seen in the figure,  $\|\hat{\omega}_2\|$  exceeds the detection threshold at 60 seconds. Since the norm of online approximator outputs for other subsystems do not reach the detection threshold, fault is only detected in the second subsystem. Therefore, in this case there is no need for the central isolation unit.



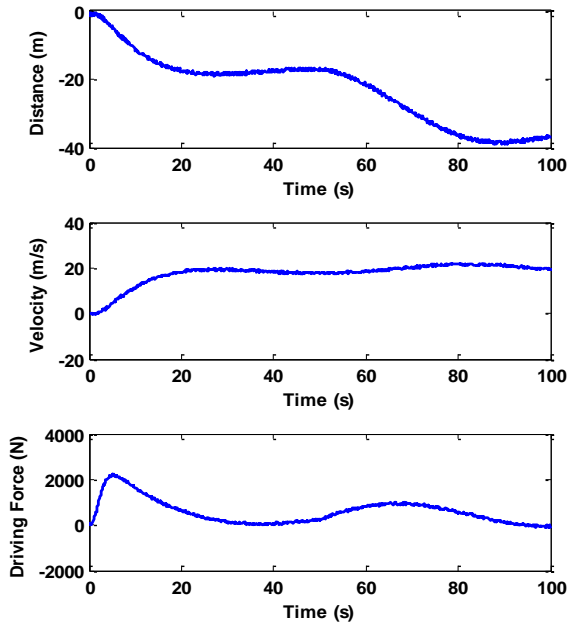


Figure 5: Estimated second vehicle states in faulty conditions

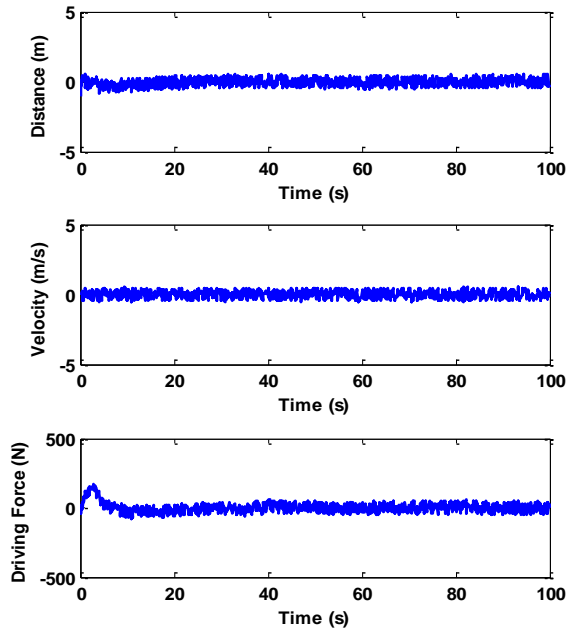


Figure 6: State estimation error for the second vehicle

Upon detection of fault, the failure prediction starts by using equation (14) with the failure limits mentioned in section 3.3. Even though the actuator fault that is initiated in subsystem 2, does not make the velocity and driving force reach their failure limits, it does affect the distance between the second and third vehicle. As illustrated in Figure 8, at time 124 seconds, a collision occurs between the second and third vehicle (Note that the inter-vehicle distances  $\psi_i$  are negative based on the definition and collision occurs when distance

reaches zero). The failure prediction algorithm continuously estimates the time-to-failure (time-to collision) after detection of fault. The result is shown in Figure 9. Time-to-failure estimation is not accurate in the first few seconds after the detection, but it is almost accurate after that, which shows the effectiveness of the prediction scheme.

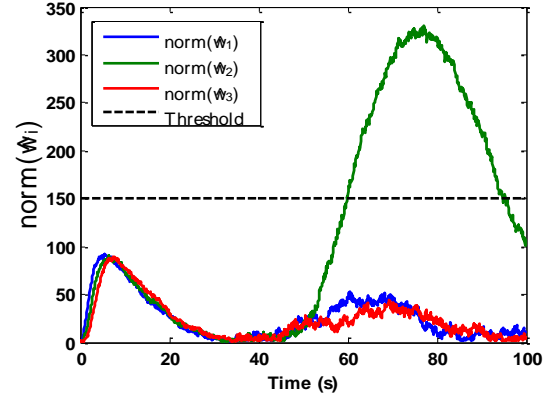


Figure 7: Norm of online approximators in the observers of the first three vehicles and the detection threshold

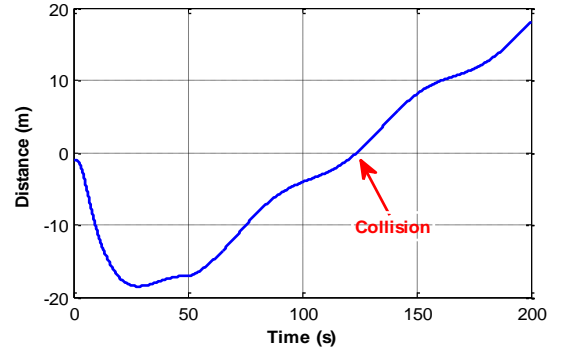


Figure 8: Distance between the second and third vehicles

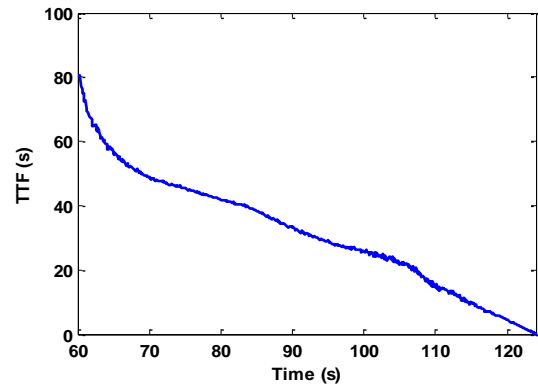


Figure 9: Time-to-failure

Next, a comparison between the proposed method and two other methods of model-based fault detection is performed by

using metrics provided by (Feldman, Kurtoglu, Narasimhan, Poll, Garcia, de Kleer, Kuhn, & van Gemund, 2010). One of the two other methods a centralized observer-based fault detection scheme (with a single observer designed for the entire system of 100 vehicles) and the other one is a semi-decentralized scheme (Stankovic et al., 2010) (which has a decentralized observer structure and assumes interconnections are known and estimated system states are transmitted to neighboring subsystems). Simulations were performed 500 times with each method. 5 different faults (multiplicative fault on sensor or actuator, bias faults on sensor or actuator, or component fault simulated by an increase in friction constant) with random magnitudes and growth rates were injected in the system at random times and in randomly selected vehicles. Only one fault occurs in each one of the simulations. The number of false positives, false negatives and average detection time is recorded for each method and summarized in Table 1. Note that false positives are not taken into account in calculation of average detection time.

The table clearly shows that both the semi-decentralized method and the proposed decentralized method outperform the centralized method in all three fault detection metrics, which is due to the fact that the centralized method is less sensitive to a single fault in one subsystem as it takes a longer time for the entire system to be affected by a fault in a single subsystem. A comparison between the results of the semi-decentralized method and the proposed decentralized method shows a small difference between their effectiveness, although the semi-decentralized method has a slightly better performance. However, the semi-decentralized method requires estimated states in each vehicle to be transmitted to the neighboring vehicles, while the proposed decentralized fault detection method does not require that. Therefore, even though the semi-decentralized method has a better performance in simulations, it will not be as efficient in practice, not only because continuous transmission of data between subsystems is not always possible, but also due to problems like delay and packet loss that can occur in transmissions and downgrade the fault detection performance.

Method	Number of false positives	Number of false negatives	Average detection time
Centralized	29	78	18.38s
Semi-decentralized	13	21	9.67s
Proposed Decentralized	17	20	11.85s

Table 1: Comparison of fault detection results

#### 4.2. Five-Tank System

In this section a five-tank water system (Ferrari et al., 2009), which is shown in Figure 10, is considered to verify the proposed decentralized fault diagnosis scheme. This

system has two input pumps with five connected water tanks. There are many different ways to decompose the system into smaller subsystems. For instance, each tank can be considered as one subsystem, or one subsystem can include tanks 1 and 2 while the other subsystem includes the three tanks. The decomposition shown in Figure 10 is one of the possible ways, where subsystem 1 includes tanks 1, 2, and 3, and subsystem 2 includes tanks 3, 4, and 5. This kind of overlapping decomposition for the five-tank system was introduced in (Ferrari et al., 2009) and it is intentionally selected in this paper to show that subsystems can be overlapping, which means they can share one or more states.

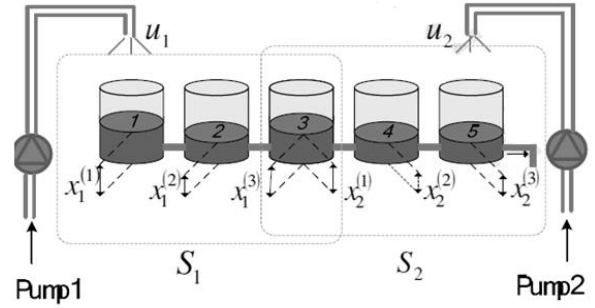


Figure 10: Five tank benchmarking system.

The system dynamics are described by (Ferrari et al., 2009)

$$\begin{aligned}
 x_1^{(1)}(k+1) &= \frac{T}{A} \left( u_1 - cs \cdot \text{sign} \left( x_1^{(1)}(k) - x_1^{(2)}(k) \right) \right. \\
 &\quad \left. \cdot \sqrt{2g \left| x_1^{(1)}(k) - x_1^{(2)}(k) \right|} \right) + \eta_1 + x_1^{(1)}(k), \\
 x_1^{(2)}(k+1) &= \frac{T}{A} \left( cs \cdot \text{sign} \left( x_1^{(1)}(k) - x_1^{(2)}(k) \right) \sqrt{2g \left| x_1^{(1)}(k) - x_1^{(2)}(k) \right|} \right. \\
 &\quad \left. - cs \cdot \text{sign} \left( x_1^{(2)}(k) - x_1^{(3)}(k) \right) \cdot \sqrt{2g \left| x_1^{(2)}(k) - x_1^{(3)}(k) \right|} \right) + \eta_2 + x_1^{(2)}(k), \\
 x_1^{(3)}(k+1) &= \frac{T}{A} \left( cs \cdot \text{sign} \left( x_1^{(2)}(k) - x_1^{(3)}(k) \right) \sqrt{2g \left| x_1^{(2)}(k) - x_1^{(3)}(k) \right|} \right. \\
 &\quad \left. - cs \cdot \text{sign} \left( x_1^{(3)}(k) - x_2^{(2)}(k) \right) \cdot \sqrt{2g \left| x_1^{(3)}(k) - x_2^{(2)}(k) \right|} \right) + \eta_3 + x_1^{(3)}(k), \\
 x_2^{(2)}(k+1) &= \frac{T}{A} \left( cs \cdot \text{sign} \left( x_1^{(3)}(k) - x_2^{(2)}(k) \right) \sqrt{2g \left| x_1^{(3)}(k) - x_2^{(2)}(k) \right|} \right. \\
 &\quad \left. - cs \cdot \text{sign} \left( x_2^{(2)}(k) - x_2^{(3)}(k) \right) \cdot \sqrt{2g \left| x_2^{(2)}(k) - x_2^{(3)}(k) \right|} \right) + \eta_4 + x_2^{(2)}(k), \\
 x_2^{(3)}(k+1) &= \frac{T}{A} \left( u_2 + cs \cdot \text{sign} \left( x_2^{(2)}(k) - x_2^{(3)}(k) \right) \right. \\
 &\quad \left. \cdot \sqrt{2g \left| x_2^{(2)}(k) - x_2^{(3)}(k) \right|} - cs \cdot \sqrt{2g x_2^{(3)}(k)} \right) + \eta_5 + x_2^{(3)}(k).
 \end{aligned}$$

where  $x_1(k) = [x_1^{(1)}(k), x_1^{(2)}(k), x_1^{(3)}(k)]^T$  is the first subsystem state vector,  $x_2(k) = [x_2^{(1)}(k), x_2^{(2)}(k), x_2^{(3)}(k)]^T$  is the second subsystem state vector,  $T$  is the sampling time

chosen to be 0.1 seconds,  $A = 0.0154 \text{ m}^2$  is the cross section of the tanks,  $s = 5 \times 10^{-5} \text{ m}^2$  is the cross section of the connecting pipes,  $c = 1$  is the outflow coefficient, and  $g = 9.8 \text{ m/s}^2$  is the standard gravity. Note that the two subsystems share one of the states, i.e.  $x_1^{(3)} = x_2^{(1)}$ . Moreover,  $\eta(x(k)) = [\eta_1(k) \ \eta_2(k) \ \eta_3(k) \ \eta_4(k) \ \eta_5(k)]^T$  represents the modeling uncertainty and is defined by

$$\eta = [10^{-2} \sin(0.7k) 10^{-2} \cos(0.8k) 10^{-2} \cos(0.5k) \\ 10^{-2} \sin(0.6k) \ 10^{-2} \cos(0.7k)]^T .$$

An incipient actuator fault in pump 1 (located in subsystem 1) is seeded at time  $t_0 = 50\text{s}$ . The dynamics of actuator fault in subsystem 1 is described by

$$\Pi(k - k_0) h_1(x_1(k), u_1(k)) \\ = \left[ \left( 1 - e^{-(k-k_0)} \right) \left( \frac{\text{sgn}[k - k_0] + 1}{2} \right) \frac{9T}{A} u_1(k), 0, 0 \right]^T .$$

Subsystem observers are designed in the form of equation (4). Online approximators  $\hat{\omega}_1$  and  $\hat{\omega}_2$  are both made up of 7-input 3-output neural networks which consists of 8 basis functions. The basis functions are sigmoid type and they satisfy  $\phi_{i_{max}} < 0.5$ . The inputs are local states at current and next time instants and the local input. Moreover, the parameter update law in (6) is used to update the neural network parameters (weights). The estimator and adaptive law parameters are taken as  $\alpha_i = 0.1$ ,  $\gamma_i = 10^{-5}$ , and  $\lambda = 0.01$ . The bounds on the uncertainty and interconnection terms are  $\eta_{i_M} = 0.029$  and  $g_{i_M} = 0.022$ . In order to calculate the detection thresholds from (10), the maximum neural network approximation errors  $\varepsilon_{i_M}$  are required. Unless the interconnection term is represented as a linear function of states and inputs,  $\varepsilon_{i_M}$  cannot be found analytically. However, the approximation error in healthy conditions is definitely less than the upper bound on interconnection term. Thus, we will replace  $\varepsilon_{i_M}$  by  $g_{i_M}$ . By using these parameter values in (10), the detection thresholds are calculated as  $\rho_i = 0.09$ .

As mentioned previously, the OLADs are always online to learn the interconnection dynamics in all subsystems. After the occurrence of fault, the OLAD in faulty subsystem will also approximate the fault dynamics. Norms of the outputs of both OLADs are plotted along with the detection threshold in Figure 11. Since the fault in subsystem 1 affects the local states as well as the interconnection terms, the output of both OLADs increase after occurrence of fault. However, the growth rate of the output of OLAD in subsystem 1 (where the fault is initiated) is significantly higher than the growth rate of the output of OLAD in subsystem 2. Thus the fault is detected first in subsystem 1.

Local residuals, which are generated by comparing the actual and estimated subsystem states, are mainly used for updating the NN weights. The norm of local residual is plotted in Figure 12 for both subsystems. Residuals are small and bounded both before and after the fault, which shows the boundedness of the state estimation errors due to successful

estimation of unknown dynamics by the stable weight update laws.

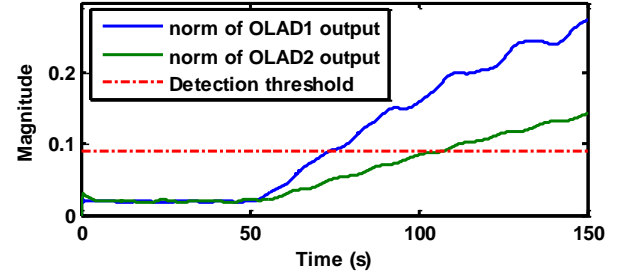


Figure 11: OLAD outputs and detection threshold.

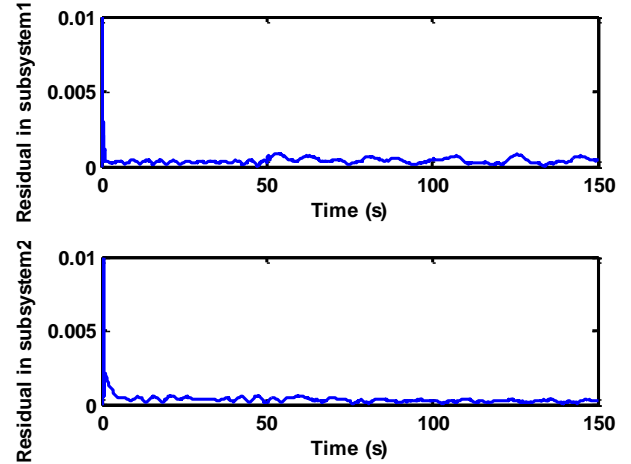


Figure 12: Residuals in subsystem 1 and subsystem 2.

The OLAD in subsystem 1 is used to estimate the vector function  $\omega_1(\cdot)$  which is the summation of interconnection term and the fault function in this subsystem. Two of the outputs of the OLAD along with their true values are shown in Figure 13. Since the fault is in tank 1 and the state equation describing tank 1 does not include an interconnection term, the first element of  $\omega_1$  (denoted by  $\omega_{11}$ ) only corresponds to the fault function. Therefore, its estimation  $\hat{\omega}_{11}$ , which is shown in Figure 13(a), reflects the approximated fault function. On the other hand, since the interconnection term appears in the state equation of the third tank and the fault is not directly affecting this tank, the third element of  $\omega_1$  (denoted by  $\omega_{13}$ ) only corresponds to the interconnection term. Therefore, its estimation  $\hat{\omega}_{13}$ , which is shown in Figure 13(b), reflects the approximated interconnection function.

The OLAD reasonably tracks the unknown vector function  $\omega_1(\cdot)$  which results in a good estimation of system states under faulty condition and allows the estimation of time-to-failure. Figure 14 shows the estimated TTF for both subsystems. The TTF is calculated for each state based on the proposed algorithm, and then the subsystem time-to-failure is obtained by taking the minimum among estimated TTF for all states of the corresponding subsystem. The TTF of subsystem 1 approaches zero faster than subsystem 2, because the fault is seeded in subsystem 1 and it has an attenuated and delayed

effect on second subsystem. Nevertheless, the entire system should be stopped before the TTF in any subsystem reaches zero. In this example, the operation of system is unsafe after  $t=100.2s$  where TTF for subsystem 1 reaches zero.

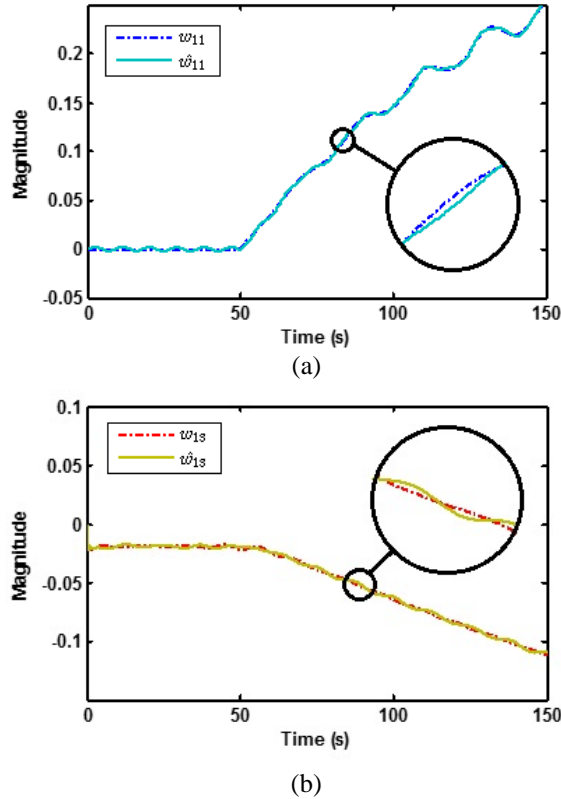


Figure 13: (a) Actual and estimated magnitude of  $\omega_{11}$ , (b) actual and estimated magnitude of  $\omega_{13}$

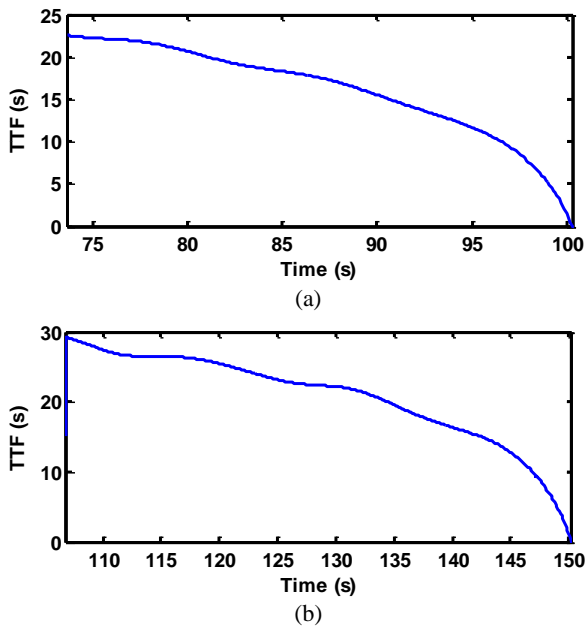


Figure 14: (a) Estimated TTF for first subsystem, (b) Estimated TTF for second subsystem.

## 5. CONCLUSIONS

The proposed decentralized fault prognosis scheme renders satisfactory performance by only using the local subsystem state vector at each LFD. A fault can be detected in all the subsystems that are significantly affected. Upon detection in each subsystem, the TTF can be predicted by using the estimated state dynamics driven by the fault approximation. The fault detection and failure prediction methods are purely decentralized and are independent of the isolation scheme that requires some data transmission.

In contrast with centralized diagnosis methods, the proposed decentralized scheme does not require transmission of large amounts of data between subsystems which saves cost and avoids transmission errors. Moreover, multiple local fault detectors increase the reliability of fault detection due to multiple layers of fault detectors. No priori offline training or fault data is necessary in order to detect or isolate faults. Hence, this scheme can save both time and cost while it is easily implementable on embedded system. The only drawbacks of the proposed scheme are the requirement for measurement of all states and the centralized isolation unit. Therefore, the future work in this topic includes development of a decentralized isolation scheme and extension of the proposed method to systems where some states are not measured.

## REFERENCES

- Bernieri, A., D'Apuzzo, M., Sansone, L., & Savastano, M. (1994). A neural network approach for identification and fault diagnosis on dynamic systems. *Instrumentation and Measurement, IEEE Transactions on*, 43(6), 867-873. doi: 10.1109/19.368083
- Blake, D., & Brown, M. (2007). *Simultaneous, Multiplicative Actuator and Sensor Fault Estimation using Fuzzy Observers*. Paper presented at the Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International.
- Boskovic, J. D., & Mehra, R. K. (2002). *A decentralized scheme for accommodation of multiple simultaneous actuator failures*. Paper presented at the American Control Conference, 2002. Proceedings of the 2002.
- Bregon, A., Daigle, M., Roychoudhury, I., Biswas, G., Koutsoukos, X., & Pulido, B. (2014). An event-based distributed diagnosis framework using structural model decomposition. *Artificial Intelligence*, 210, 1-35.
- Caccavale, F., & Villani, L. (2004, June 30 2004-July 2 2004). *An adaptive observer for fault diagnosis in nonlinear discrete-time systems*. Paper presented at the American Control Conference, 2004. Proceedings of the 2004.
- Dash, S., & Venkatasubramanian, V. (2000). Challenges in the industrial applications of fault diagnostic systems. *Computers & Chemical Engineering*, 24(2-7), 785-791. doi: 10.1016/S0098-1354(00)00374-4
- Demetriou, M. A., & Polycarpou, M. M. (1998). Incipient fault diagnosis of dynamical systems using online approximators. *Automatic Control, IEEE Transactions on*, 43(11), 1612-1617. doi: 10.1109/9.728881
- Daigle, M., & Goebel, K. (2009). Model-based prognostics with fixed-lag particle filters. *Annual Conference of the Prognostics and Health Management Society*, San Diego, CA.

- Feldman, A., Kurtoglu, T., Narasimhan, S., Poll, S., Garcia, D., de Kleer, J., Kuhn, L., & van Gemund, A. (2010). Empirical evaluation of diagnostic algorithm performance using a generic framework. *International Journal of Prognostics and Health Management*, 1(2), 1-28.
- Ferdowsi, H., & Jagannathan, S. (2013). A unified model-based fault diagnosis scheme for non-linear discrete-time systems with additive and multiplicative faults. *Transactions of the Institute of Measurement and Control*, 35(6), 742-752. doi: 10.1177/0142331212473141
- Ferdowsi, H., & Jagannathan, S. (2017). Decentralized fault tolerant control of a class of nonlinear interconnected systems. *International Journal of Control, Automation and Systems*, 15(2), 527-536.
- Ferdowsi, H., Raja, D. L., & Jagannathan, S. (2012a). A decentralized fault prognosis scheme for nonlinear interconnected discrete-time systems. 2012 American Control Conference (ACC), 5900-5905.
- Ferdowsi, H., Raja, D. L., & Jagannathan, S. (2012b). A decentralized fault detection and prediction scheme for nonlinear interconnected continuous-time systems. Neural Networks (IJCNN), The 2012 International Joint Conference on, 1-7.
- Ferrari, R., Parisini, T., & Polycarpou, M. M. (2009). Distributed Fault Diagnosis With Overlapping Decompositions: An Adaptive Approximation Approach. *Automatic Control, IEEE Transactions on*, 54(4), 794-799.
- Golub, G. H., & Loan, C. F. V. (1996). *Matrix computations (3rd ed.)*: Johns Hopkins University Press.
- Huang, S. N., Tan, K. K., & Lee, T. H. (2005). Decentralized control of a class of large-scale nonlinear systems using neural networks. *Automatica*, 41(9), 1645-1649. doi: 10.1016/j.automatica.2005.02.010
- Huang, S. N., Tan, K. K., & Lee, T. H. (2006). Nonlinear adaptive control of interconnected systems using neural networks. *Neural Networks, IEEE Transactions on*, 17(1), 243-246. doi: 10.1109/tnn.2005.857948
- Isermann, R. (2005). Model-based fault-detection and diagnosis – status and applications. *Annual Reviews in Control*, 29(1), 71-85. doi: 10.1016/j.arcontrol.2004.12.002
- Jagannathan, S. (2006). *Neural Network Control of Nonlinear Discrete-time Systems*. NY: CRC publications.
- Kulkarni, C., Daigle, M., Gorospe, G., & Goebel, K. (2014). Validation of model-based prognostics for pneumatic valves in a demonstration testbed. *Annual Conference of the Prognostics and Health Management Society*, 76-85.
- Luo, J., Namburu, M., Pattipati, K., Qiao, L., Kawamoto, M., & Chigusa, S. A. C. S. (2003). Model-based prognostic techniques [maintenance applications]. In *AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference. Proceedings*, 330-340.
- Maki, Y., & Loparo, K. A. (1997). A neural-network approach to fault detection and diagnosis in industrial processes. *Control Systems Technology, IEEE Transactions on*, 5(6), 529-541. doi: 10.1109/87.641399
- Patton, R. J., Chen, J., & Lopez-Toribio, C. J. (1998, 1998). *Fuzzy observers for nonlinear dynamic systems fault diagnosis*. Paper presented at the Decision and Control, 1998. Proceedings of the 37th IEEE Conference on.
- SamPATH, M., Sengupta, R., Lafortune, S., Sinnamohideen, K., & Teneketzis, D. (1995). Diagnosability of discrete-event systems. *IEEE Transactions on Automatic Control*, 40(9), 1555-1575.
- Stankovic, S., Ilic, N., Djurovic, Z., Stankovic, M., & Johansson, K. H. (2010, 6-8 Oct. 2010). *Consensus based overlapping decentralized fault detection and isolation*. Paper presented at the Control and Fault-Tolerant Systems (SysTol), 2010 Conference on.
- Thumati, B. T., & Jagannathan, S. (2010). A Model-Based Fault-Detection and Prediction Scheme for Nonlinear Multivariable Discrete-Time Systems With Asymptotic Stability Guarantees. *Neural Networks, IEEE Transactions on*, 21(3), 404-423. doi: 10.1109/tnn.2009.2037498
- Wang, H., & Daley, S. (1996). Actuator fault diagnosis: an adaptive observer-based technique. *Automatic Control, IEEE Transactions on*, 41(7), 1073-1078. doi: 10.1109/9.508919
- Yan, X. G., & Edwards, C. (2008). Robust decentralized actuator fault detection and estimation for large-scale systems using a sliding mode observer. *International Journal of control*, 81(4), 591-606.
- Zhang, J., & Morris, A. J. (1994). On-line process fault diagnosis using fuzzy neural networks. *Intelligent Systems Engineering*, 3(1), 37-47.

## APPENDIX

**Proof of Theorem 1:** Consider the following Lyapunov function candidate

$$V = \frac{1}{8\phi_{\max}^2} e_i^T(k-1)e_i(k-1) + \text{tr}\{\tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1)\}. \quad (\text{A.1})$$

The Lyapunov function is deliberately selected at time  $k-1$ , because the observer is one time step behind the actual system and its output is not available at time  $k$ . In other words,  $k-1$  is the current time instant for the observer. The first difference of the Lyapunov function is given by

$$\Delta V = \underbrace{\frac{1}{8\phi_{\max}^2} (e_i^T(k)e_i(k) - e_i^T(k-1)e_i(k-1))}_{\Delta V_1} + \underbrace{\text{tr}\{\tilde{\theta}_i^T(k)\tilde{\theta}_i(k) - \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1)\}}_{\Delta V_2}. \quad (\text{A.2})$$

Substitute  $e_i(k)$  from the local error dynamics (9), in  $\Delta V_1$  to get

$$\begin{aligned} \Delta V_1 = & \frac{1}{8\phi_{\max}^2} \left[ \lambda e_i(k-1) + \eta_i(x_i(k-1), u_i(k-1)) + \varepsilon_i(k-1) \right. \\ & \left. + \tilde{\theta}_i^T(k-1)\phi_i(x_i(k), x_i(k-1), u_i(k-1)) \right]^T \\ & \left[ \lambda e_i(k-1) + \eta_i(x_i(k-1), u_i(k-1)) + \varepsilon_i(k-1) \right. \\ & \left. + \tilde{\theta}_i^T(k-1)\phi_i(x_i(k), x_i(k-1), u_i(k-1)) \right] \\ & - \frac{1}{8\phi_{\max}^2} e_i^T(k-1)e_i(k-1). \end{aligned}$$

By using the Cauchy-Schwarz inequality ( $(s_1+s_2+\dots+s_n)^T(s_1+s_2+\dots+s_n) \leq n(s_1^T s_1 + s_2^T s_2 + \dots + s_n^T s_n)$ ) we arrive at

$$\begin{aligned}
\Delta V_1 \leq & \frac{1}{2\phi_{i\max}^2} \lambda^2 e_i^T(k-1)e_i(k-1) + \frac{1}{2\phi_{i\max}^2} \varepsilon_i^T(k-1)\varepsilon_i(k-1) \\
& + \frac{1}{2\phi_{i\max}^2} \eta_i^T(x_i(k-1), u_i(k-1))\eta_i(x_i(k-1), u_i(k-1)) \\
& + \frac{1}{2\phi_{i\max}^2} \phi_i^T(k-1)\tilde{\theta}_i(k-1)\tilde{\theta}_i^T(k-1)\phi_i(k-1) \\
& - \frac{1}{8\phi_{i\max}^2} e_i^T(k-1)e_i(k-1). \tag{A.3}
\end{aligned}$$

Now substitute  $\hat{\theta}_i(k)$  from (6), in  $\Delta V_2$

$$\begin{aligned}
\Delta V_2 = & \text{tr} \left\{ \left[ (1-\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|) \tilde{\theta}_i(k-1) \right. \right. \\
& \left. \left. + \gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| \theta_i - \alpha_i \phi_i(k-1)e_i^T(k-1) \right]^T \right. \\
& \left. \cdot \left[ (1-\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|) \tilde{\theta}_i(k-1) \right. \right. \\
& \left. \left. + \gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| \theta_i - \alpha_i \phi_i(k-1)e_i^T(k-1) \right] \right. \\
& \left. - \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right\}
\end{aligned}$$

By applying the Cauchy-Schwarz inequality we get

$$\begin{aligned}
\Delta V_2 \leq & 3\text{tr} \left\{ \alpha_i^2 e_i(k-1)\phi_i^T(k-1)\phi_i(k-1)e_i^T(k-1) \right. \\
& \left. + \gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \theta_i^T(k-1)\theta_i(k-1) \right. \\
& \left. + \left( (1-\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|) \right)^2 \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right\} \\
& - \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \\
\leq & \text{tr} \left\{ 2\tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right. \\
& \left. - 6\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right. \\
& \left. + 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right\} \\
& + \text{tr} \left\{ 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \theta_i^T(k-1)\theta_i(k-1) \right\} \\
& + 3\alpha_i^2 e_i^T(k-1)e_i(k-1)\phi_i^T(k-1)\phi_i(k-1). \tag{A.4}
\end{aligned}$$

By combining  $\Delta V_1$  and  $\Delta V_2$  from (A.3) and (A.4) we get

$$\begin{aligned}
\Delta V = & \Delta V_1 + \Delta V_2 \\
\leq & \frac{1}{2\phi_{i\max}^2} \lambda^2 e_i^T(k-1)e_i(k-1) + \frac{1}{2\phi_{i\max}^2} \varepsilon_i^T(k-1)\varepsilon_i(k-1) \\
& + \frac{1}{2\phi_{i\max}^2} \eta_i^T(x_i(k-1), u_i(k-1))\eta_i(x_i(k-1), u_i(k-1)) \\
& + \frac{1}{2\phi_{i\max}^2} \phi_i^T(k-1)\tilde{\theta}_i(k-1)\tilde{\theta}_i^T(k-1)\phi_i(k-1) \\
& - \frac{1}{8\phi_{i\max}^2} e_i^T(k-1)e_i(k-1) \\
& + \text{tr} \left\{ 2\tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right. \\
& \left. - 6\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right. \\
& \left. + 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \tilde{\theta}_i^T(k-1)\tilde{\theta}_i(k-1) \right\} \\
& + \text{tr} \left\{ 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \theta_i^T(k-1)\theta_i(k-1) \right\} \\
& + 3\alpha_i^2 e_i^T(k-1)e_i(k-1)\phi_i^T(k-1)\phi_i(k-1).
\end{aligned}$$

Taking the Frobenius norm (Golub & Loan, 1996), and using the result of assumptions 1 and 2, we get

$$\begin{aligned}
\Delta V \leq & -\left( \frac{1}{8\phi_{i\max}^2} - \frac{\lambda^2}{2\phi_{i\max}^2} \right) \|e_i(k-1)\|^2 + 3\alpha_i^2 \phi_{i\max}^2 \|e_i(k-1)\|^2 \\
& - \left( 6\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| - 2 - \frac{1}{2} \right. \\
& \left. - 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \right) \|\tilde{\theta}_i(k-1)\|^2 \\
& + \frac{\eta_{iM}^2}{2\phi_{i\max}^2} + \frac{\varepsilon_{iM}^2}{2\phi_{i\max}^2} + 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \theta_{i\max}^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Delta V \leq & -\left( \frac{1}{8\phi_{i\max}^2} - \frac{\lambda^2}{2\phi_{i\max}^2} - 3\alpha_i^2 \phi_{i\max}^2 \right) \|e_i(k-1)\|^2 \\
& - \left( 6\gamma_i \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\| - 2.5 \right. \\
& \left. - 3\gamma_i^2 \|I - \alpha_i \phi_i(k-1)\phi_i^T(k-1)\|^2 \right) \|\tilde{\theta}_i(k-1)\|^2 \\
& + \frac{\eta_{iM}^2 + \varepsilon_{iM}^2}{2\phi_{i\max}^2} + 3\gamma_i^2 (1 + \alpha_i \phi_{i\max}^2)^2 \theta_{i\max}^2. \tag{A.5}
\end{aligned}$$

In order to prove the uniform ultimate boundedness of  $e_i$  and  $\theta_i$ , the following conditions must be satisfied

$$\frac{1-4\lambda^2}{8\phi_{i\max}^2} - 3\alpha_i^2 \phi_{i\max}^2 > 0,$$

$$-3\gamma_i^2 \|I - \alpha_i \phi_i \phi_i^T\|^2 + 6\gamma_i \|I - \alpha_i \phi_i \phi_i^T\| - 2.5 > 0.$$

The first condition is satisfied if  $\lambda$  and  $\alpha_i$  are selected such that  $|\lambda| < 0.5$  and  $\alpha_i < \sqrt{(1-4\lambda^2)/24\phi_{i\max}^4}$ . The second condition is equivalent to  $0.6 < \gamma_i \|I - \alpha_i \phi_i \phi_i^T\| < 1.4$  and since  $\gamma_i \leq \gamma_i \|I - \alpha_i \phi_i \phi_i^T\| \leq \gamma_i (1 + \alpha_i \phi_{i\max}^2)$  always

holds, the second condition is satisfied if  $\gamma_i$  is selected such that  $0.6 < \gamma_i < \frac{1.4}{1 + \alpha_i \phi_{i_{\max}}^2}$ .

Based on the above discussion, when the user-defined parameters are selected as stated in theorem 1, the first difference of the Lyapunov function will be less than zero, if at least one of the following conditions is satisfied

$$\|e_i\| > \sqrt{D/C_1} \quad \text{or} \quad \|\tilde{\theta}_i\| > \sqrt{D/C_2}, \quad (\text{A.6})$$

where

$$C_1 = (1 - 4\lambda^2) / 8\phi_{i_{\max}}^2 - 3\alpha_i^2 \phi_{i_{\max}}^2,$$

$$C_2 = -3\gamma_i^2 \|I - \alpha_i \phi_i \phi_i^T\|^2 + 6\gamma_i \|I - \alpha_i \phi_i \phi_i^T\| - 2.5,$$

$$D = (\eta_{i_M}^2 + \varepsilon_{i_M}^2) / 2\phi_{i_{\max}}^2 + 3\gamma_i^2 (1 + \alpha_i \phi_{i_{\max}}^2)^2 \theta_{i_{\max}}^2.$$

Therefore, with the appropriate choice of design parameters, the local FD residual,  $e_i(k)$ , and the parameter estimation error  $\tilde{\theta}_i(k)$ , will be uniformly ultimately bounded with the bounds given in (A.6).

**Proof of Theorem 2:** Consider the output of local OLAD in subsystem  $i$

$$\begin{aligned} \hat{\omega}_i(k) &= \hat{\theta}_i^T(k) \phi_i(x_i(k+1), x_i(k), u_i(k)) \\ &= \theta_i^T(k) \phi_i(x_i(k+1), x_i(k), u_i(k)) \\ &\quad - \tilde{\theta}_i^T(k) \phi_i(x_i(k+1), x_i(k), u_i(k)) \\ &= \omega_i(k) - \varepsilon_i(k) - \tilde{\theta}_i^T(k) \phi_i(x_i(k+1), x_i(k), u_i(k)). \end{aligned}$$

By taking Frobenius norm and using Assumptions 1, 3, and 4 as well as the result of theorem 1, we get

$$\begin{aligned} \|\hat{\omega}_i(k)\| &\leq \|g_i(x(k), u_i(k))\| + \|\varepsilon_i(k)\| + \phi_{i_{\max}} \|\tilde{\theta}(k)\| \\ &\leq g_{i_M} + \varepsilon_{i_M} + \phi_{i_{\max}} \sqrt{D/C_2}. \end{aligned} \quad (\text{A.7})$$

Therefore, if the detection threshold is selected as in (8), then no fault is detected as long as the system is working under healthy operating conditions.

To find the detectability condition, the output of OLAD in the faulty subsystem is utilized

$$\begin{aligned} \|\hat{\omega}_i(k_d)\| &\geq \|h_i(x_i(k_d), u_i(k_d))\| - \|g_i(x(k_d), u_i(k_d))\| \\ &\quad - \|\varepsilon_i(k_d)\| - \phi_{i_{\max}} \|\tilde{\theta}(k_d)\| \\ &\geq \|h_i(x_i(k_d), u_i(k_d))\| - g_{i_M} - \varepsilon_{i_M} - \phi_{i_{\max}} \sqrt{D/C_2}. \end{aligned}$$

Therefore, if there exist a time instant  $k_d$  at which the following condition is satisfied

$$\|h_i(x_i(k_d), u_i(k_d))\| - g_{i_M} - \varepsilon_{i_M} - \phi_{i_{\max}} \sqrt{D/C_2} \geq \rho_i, \quad (\text{A.8})$$

or equivalently

$$\|h_i(x_i(k_d), u_i(k_d))\| \geq 2\rho_i - q_i,$$

then the fault will be detected in the faulty subsystem.

**Proof of Theorem 3:** Consider the observer dynamics in (4) rewritten as

$$\hat{x}_i(k) = \lambda \hat{x}_i(k-1) + b s_i(k-1), \quad (\text{A.9})$$

where  $b=1$  and  $s_i$ , defined in (15), acts as the input to the linear system of (A.9). By assuming that the fault is detected at time  $k_d$ , the response to this set of linear state space equations at time  $k_f > k_d$  is given by

$$\hat{x}_i(k_f) = \lambda^{k_f - k_d + 1} \hat{x}_i(k_d) + \sum_{l=k_d}^{k_f-1} \lambda^{k_f - l - 1} b s_i(l). \quad (\text{A.10})$$

If we assume that  $s_i(l) = s_i(k-1)$  for  $k-1 \leq l \leq k_f$  (which is reasonable, since the fault is assumed to be incipient type), we will have

$$\begin{aligned} \hat{x}_i(k_f) &= \lambda^{k_f - k_d + 1} \hat{x}_i(k_d) + b s_i(k-1) \sum_{l=k_d}^{k_f-1} \lambda^{k_f - l - 1} \\ &= \lambda^{k_f - k_d + 1} \hat{x}_i(k_d) + b s_i(k-1) \frac{1 - \lambda^{k_f - k_d + 1}}{1 - \lambda}. \end{aligned}$$

Now suppose that  $k_{f_{i,j}}$  is the time when the  $j^{\text{th}}$  state of subsystem  $i$ , reaches its failure threshold, i.e.  $\hat{x}_{i,j}(k_{f_{i,j}}) = \bar{x}_{i,jM}$ .

$$\begin{aligned} x_{i,jM} &= \lambda^{k_{f_{i,j}} - k_d + 1} \hat{x}_{i,j}(k_d) + s_{i,j}(k-1) \frac{1 - \lambda^{k_{f_{i,j}} - k_d + 1}}{1 - \lambda} \\ &\Rightarrow (1 - \lambda) x_{i,jM} - s_{i,j}(k-1) \\ &= \lambda \lambda^{k_{f_{i,j}} - k} \left( (1 - \lambda) \hat{x}_{i,j}(k-1) - s_{i,j}(k-1) \right) \\ &\Rightarrow \frac{(1 - \lambda) x_{i,jM} - s_{i,j}(k-1)}{\lambda (1 - \lambda) \hat{x}_{i,j}(k-1) - \lambda s_{i,j}(k-1)} = \lambda^{k_{f_{i,j}} - k}. \end{aligned}$$

Therefore, the time to failure for the  $j^{\text{th}}$  state of the  $i^{\text{th}}$  subsystem can be estimated by

$$TTF_{i,j} = k_{f_{i,j}} - k = \log_{\lambda} \left| \frac{(1 - \lambda) x_{i,jM} - s_{i,j}(k-1)}{\lambda (1 - \lambda) \hat{x}_{i,j}(k-1) - \lambda s_{i,j}(k-1)} \right|.$$