Improving Multiple Fault Diagnosability using Possible Conflicts [⋆]

Matthew Daigle [*] **Anibal Bregon** [**] **Gautam Biswas** [***]
Xenofon Koutsoukos [***] **Belarmino Pulido** [**]

[*] *NASA Ames Research Center, Moffett Field, CA, 94035, USA*
(e-mail: matthew.j.daigle@nasa.gov).
[**] *Departamento de Informatica, University of Valladolid, Valladolid,*
47011, Spain (e-mail: anibal@infor.uva.es)
[***] *Institute for Software Integrated Systems, Dept. of EECS,*
Vanderbilt University, Nashville, TN, 37235, USA

Abstract: Multiple fault diagnosis is a difficult problem for dynamic systems. Due to fault masking, compensation, and relative time of fault occurrence, multiple faults can manifest in many different ways as observable fault signature sequences. This decreases diagnosability of multiple faults, and therefore leads to a loss in effectiveness of the fault isolation step. We develop a qualitative, event-based, multiple fault isolation framework, and derive several notions of multiple fault diagnosability. We show that using Possible Conflicts, a model decomposition technique that decouples faults from residuals, we can significantly improve the diagnosability of multiple faults compared to an approach using a single global model. We demonstrate these concepts and provide results using a multi-tank system as a case study.

Keywords: Diagnosability, multiple fault diagnosis, structural analysis, discrete-event systems

1. INTRODUCTION

Multiple simultaneous faults in a system add significant complexity to the fault diagnosis problem, especially in dynamic systems. Fault masking, compensation, and the relative time of fault occurrence give rise to many different ways that multiple faults can manifest in the system observations. As a result, isolating multiple faults becomes a difficult task. The larger the number of faults considered, the more possible ways their effects can interleave, making it less likely that the fault candidates can be uniquely isolated given a set of measurements.

Typically, multiple fault diagnosis (MFD) solutions apply to static systems, e.g., (de Kleer and Williams, 1987). For dynamic systems, (Dvorak and Kuipers, 1991) performs qualitative and semi-quantitative simulation to mimic the evolution of the process, changing the configuration of the model every time a fault appears. (Nyberg and Krysander, 2003) integrates FDI techniques for fault detection and DX techniques for fault isolation that can automatically handle multiple faults in dynamic systems.

Our previous work in MFD for continuous systems (Daigle et al., 2007; Daigle, 2008), based on a qualitative fault isolation (QFI) framework (Mosterman and Biswas, 1999) described how multiple faults manifest in the measurements, and provided algorithms for fault isolation. This approach was based on using residuals computed from a global model. Since faults affect all measurements that have a causal path from the fault to the measurement, fault masking can have a significant adverse impact on multiple fault diagnosability.

Using analytical redundancy relations (ARRs) approaches, diagnosability is improved by deriving relations that decouple faults from residuals, so that a single fault affects only a small set of residuals (Gertler, 1998). This decreases the possibility of masking, and, as such, should intuitively lead to improvements in multiple fault diagnosability. In this work, we explore this idea using the model decomposition approach of Possible Conflicts (PCs) (Pulido and Alonso-González, 2004), which is a dependency-compilation technique that automatically partitions the system model into minimal over-determined subsystems, based on the set of measurements and faulty components. PCs are designed to be triggered only by faults within its subsystem, thus decoupling faults from residuals.

In this paper, we develop a qualitative, event-based framework for multiple fault diagnosis that takes advantage of model decomposition. We develop several notions of multiple fault distinguishability that are applicable depending on what assumptions the user is willing to make. We define multiple fault diagnosability and provide a means to quantify it for a system. Using a tank system as a case study, we show how using residuals derived from PCs improves multiple fault diagnosability of a system, and how a combined approach using residuals derived from both the global model and the PCs further improves diagnosability.

The paper is organized as follows. Section 2 describes preliminary material. Section 3 reviews residual generation and model decomposition using PCs. Section 4 overviews the QFI framework and Section 5 describes event-based

[⋆] Anibal Bregon and Belarmino Pulido's work has been funded by the Spanish MCI DPI2008-01996 and TIN2009-11326 grants.

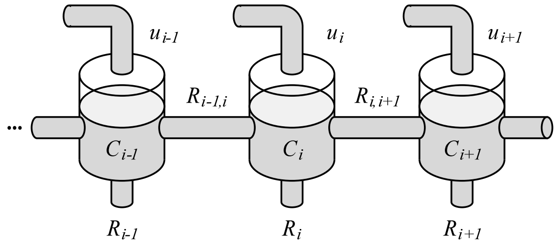


Fig. 1. Tank system schematic.

fault modeling. Section 6 establishes distinguishability and diagnosability of multiple faults within our framework. Section 7 applies the framework to a tank system case study. Section 8 concludes the paper.

2. PRELIMINARIES

We assume the system is described by

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t)) + \mathbf{v}(t)$$
$$\mathbf{y}(t) = \mathbf{h}(\mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t)) + \mathbf{w}(t),$$

where $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ is the state, $\boldsymbol{\theta}(t) \in \mathbb{R}^{n_\theta}$ is the parameter vector, $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ is the input, $\mathbf{v}(t) \in \mathbb{R}^{n_v}$ is the process noise, $\mathbf{f}$ and $\mathbf{h}$ are the state and output equations, respectively, $\mathbf{y}(t) \in \mathbb{R}^{n_y}$ is the output, and $\mathbf{w}(t) \in \mathbb{R}^{n_w}$ is the measurement noise.

We denote a measurement as m, which refers to an output variable in $\mathbf{y}$, and a measurement set as M. We consider abrupt parametric faults, with faults modeled as persistent unexpected step changes in system parameter values. We name faults by the associated parameter and the direction of change, e.g., θ^+ denotes a fault defined as an abrupt increase in the value of parameter θ. We denote a fault as f and a set of faults as F.

In MFD, a candidate is defined as a set of faults.

Definition 1. (Candidate). A *candidate* $c \subseteq F$ is a set of faults. The set of all candidates is denoted as C.

For example, the candidate $\{f_1, f_2\}$ (in shorthand, $f_1 f_2$) implies that both f_1 and f_2 have occurred.

In this paper, we focus specifically on the diagnosability of a system with multiple faults. That is, we study the distinguishability of candidates in C within our fault isolation framework. In this paper, we do not take minimality of candidates into account, i.e., we still want to be able to distinguish between candidates c_1 and c_2 even if $c_1 \subset c_2$. Working with minimal candidates in our framework is described in (Daigle et al., 2007; Daigle, 2008) and more generally in (de Kleer and Williams, 1987).

Throughout the paper, we use a multi-tank system as a running example. A number of tanks are connected serially (see Fig. 1). For tank i, u_i is the input flow, C_i is the tank capacitance, and R_i is the drain pipe resistance. For tanks i and j, R_{ij} is the connecting pipe resistance. For an n-tank system, the pressure of tank i, p_i, is described by

$$\dot{p}_i = \frac{1}{C_i}\Big(u_i + q_{i-1,i} - q_i - q_{i,i+1}\Big) + v_i,$$

where v_i is the process noise for tank i, $q_i = \frac{1}{R_i}(p_i)$ is the output flow of tank i, and $q_{i,i+1} = \frac{1}{R_{i,i+1}}(p_i - p_{i+1})$ is the flow between tanks i and $i+1$. For tank 1, $q_{0,1} = 0$, and for tank n, $q_{n,n+1} = 0$.

The complete fault set F consists of $\{C_i^-, C_i^+, R_i^-, R_i^+ : i = 1, \ldots, n\} \cup \{R_{i,i+1}^-, R_{i,i+1}^+ : i = 1, \ldots, n-1\}$. The complete measurement set M is defined as $\{p_i, q_i : i = 1, \ldots, n\} \cup \{q_{i,i+1} : i = 1, \ldots, n-1\}$. We consider single and double faults to form the candidate set C, so there are $|F| + \binom{|F|}{2} = 8 + 28 = 36$ candidates [1].

3. MODEL DECOMPOSITION

In our previous approach (Daigle et al., 2007; Daigle, 2008), a global system model was used for residual generation. An observer, based on the global model, is used to estimate the system behavior based on the set of measurements (Mosterman and Biswas, 1999). This estimate is then used to compute a residual, r, for the measurement, i.e., r is computed as the difference between an observation, y, and its predicted nominal values, $\hat{y}$, i.e., $r(t) = y(t) - \hat{y}(t)$. Therefore, we compute a residual for each measurement of the system. We denote a residual as r_m, where m is the associated measurement, and the residual set is denoted as R.

With model decomposition methods, like PCs, the global model is decomposed into a set of minimal over-determined subsystems, each with a single output (one submodel per measurement) [2] (Pulido and Alonso-González, 2004). We define residuals in the same way, only the predicted output $\hat{y}$ is computed using an observer based on the submodel computing y. The submodels are made independent of each other by using measurements as inputs to the submodels. As a result, a single fault is found in only a few submodels (ideally, one submodel), and, therefore, a fault affects only a subset of the residuals and is decoupled from the rest. Intuitively, this decoupling should improve multiple fault diagnosability. For example, if two faults do not affect any common residuals, we should be able to distinguish between the situation where only one of the faults occurs and both have occurred. With the global model approach, without such decoupling, one fault may completely mask the other, preventing distinguishability.

Applying the PCs approach to an n-tank system with $M = \{p_i : i = 1, \ldots, n\}$ we find a set of n minimal submodels. Each PC, PC_i, estimates the pressure in one of the tanks, p_i, and can be described in a general way as follows:

$$\dot{p}_i = \frac{1}{C_i}\Big(u_i + \frac{(p'_{i-1} - p_i)}{R_{i-1,i}} - \frac{(p_i)}{R_i} - \frac{(p_i - p'_{i+1})}{R_{i,i+1}}\Big),$$

where p_i is the state variable, u_i is the input to the tank, p'_{i-1} and p'_{i+1} are the measured pressures of tanks $i-1$ and $i+1$ that are used as input for the PC, and $\{C_i, R_i, R_{i,i-1}, R_{i,i+1}\}$ is the subset of (fault) parameters that affects the estimation of PC_i. For example, using the PC approach with a three-tank system with $M = \{p_1, p_2, p_3\}$ we find three PCs, each one of them estimating the pressure in one of the tanks.

[1] Note that our approach is not limited by candidate cardinality. We focus here only on single and double faults for demonstration.

[2] PCs have been demonstrated to be equivalent to other structural methods for residual generation, such as minimal ARRs (Pulido and Alonso-González, 2004).

Table 1. Fault Signatures and RMO for the Global Model of the Tank System.

Fault	r_{p_1}	r_{p_2}	r_{p_3}	Measurement Orderings
C_1^-	+-	0+	0+	$r_{p_1} \prec r_{p_2}, r_{p_1} \prec r_{p_3}, r_{p_2} \prec r_{p_3}$
R_1^+	0+	0+	0+	$r_{p_1} \prec r_{p_2}, r_{p_1} \prec r_{p_3}, r_{p_2} \prec r_{p_3}$
R_{12}^+	0+	0-	0-	$r_{p_2} \prec r_{p_3}$
C_2^-	0+	+-	0+	$r_{p_2} \prec r_{p_1}, r_{p_2} \prec r_{p_3}$
R_2^+	0+	0+	0+	$r_{p_2} \prec r_{p_1}, r_{p_2} \prec r_{p_3}$
R_{23}^+	0+	0+	0-	$r_{p_2} \prec r_{p_1}$
C_3^-	0+	0+	+-	$r_{p_2} \prec r_{p_1}, r_{p_3} \prec r_{p_1}, r_{p_3} \prec r_{p_2}$
R_3^+	0+	0+	0+	$r_{p_2} \prec r_{p_1}, r_{p_3} \prec r_{p_1}, r_{p_3} \prec r_{p_2}$

4. QUALITATIVE FAULT ISOLATION

Faults cause deviations in the measured variables from the nominal values. Residual deviations are abstracted using qualitative +, -, and 0 values to form *fault signatures* (Mosterman and Biswas, 1999). Fault signatures represent these deviations as the immediate change in magnitude and the first nonzero derivative change.

Definition 2. (Fault Signature). A *fault signature* for a fault f and residual r is the qualitative change in magnitude and slope of r caused by the occurrence of f, and is denoted by $\sigma_{f,r} \in \Sigma_{f,r}$.

Note that due to possible ambiguities in the fault signatures, $\sigma_{f,r}$ may not be unique. A fault signature is written as $s_1 s_2$, where s_1 is the qualitative magnitude change and s_2 is the qualitative slope change, e.g., -+.

We also capture the temporal order of residual deviations for a given submodel, termed *relative measurement orderings* (RMOs), based on the intuition that fault effects will manifest in some parts of the system before others (Daigle, 2008). They are computed based on analysis of the transfer functions from faults to residuals defined for measurements within a submodel.

Definition 3. (Relative Measurement Ordering). If fault f manifests in residual r_i before residual r_j, then we define a *relative measurement ordering* between r_i and r_j for fault f, denoted by $r_i \prec_f r_j$. We denote the set of all measurement orderings for f as $\Omega_{f,R}$.

Because RMOs are defined only within a given submodel, they cannot be straightforwardly computed between residuals of two different submodels because they are independent. Such RMOs will not be considered when using PCs.

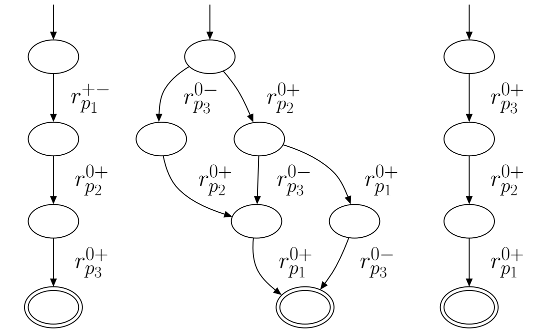
Signatures and RMOs can be computed automatically from a system model (Daigle, 2008). Table 1 shows these for the global model of a three-tank system with $F = \{C_1^-, C_2^-, C_3^-, R_1^+, R_2^+, R_3^+, R_{12}^+, R_{23}^+\}$, $M = \{p_1, p_2, p_3\}$, and $R = \{r_{p_1}, r_{p_2}, r_{p_3}\}$. Signatures derived from the PCs with residuals from the same measurements are shown in Table 2. In this case, each residual is only affected by a subset of the faults, e.g., C_1^-, causes a discontinuous increase in r_{p_1} for both approaches, followed by a smooth decrease, denoted by the signature +-. This is followed by smooth increases in residuals r_{p_2} and r_{p_3} for the global model, but no effect appears in these residuals for the PCs.

5. EVENT-BASED FAULT MODELING

Fault signatures combined with RMOs provide event-based information for diagnosis. For a given fault, the combination of all fault signatures and measurement orderings yields all the possible ways a fault can manifest in the residuals. We define each of these possibilities as a *fault trace*.

Table 2. Fault Signatures and RMO for the Set of Minimal Submodels of the Tank System.

Fault	r_{p_1}	r_{p_2}	r_{p_3}	Measurement Orderings
C_1^-	+-	00	00	∅
R_1^+	0+	00	00	∅
R_{12}^+	0+	0-	00	∅
C_2^-	00	+-	00	∅
R_2^+	00	0+	00	∅
R_{23}^+	00	0+	0-	∅
C_3^-	00	00	+-	∅
R_3^+	00	00	0+	∅



(a) $\mathcal{L}_{C_1^-,R}$ (b) $\mathcal{L}_{R_{23}^+,R}$ (c) $\mathcal{L}_{R_3^+,R}$

Fig. 2. Some fault models (obtained using the global model), where $R = \{r_{p_1}, r_{p_2}, r_{p_3}\}$.

Definition 4. (Fault Trace). A *fault trace* for a fault f over residuals R, denoted by $\lambda_{f,R}$, is a string of length $\leq |R|$ that includes, for every $r \in R$ that will deviate due to f, a fault signature $\sigma_{f,r}$, such that the sequence of fault signatures satisfies $\Omega_{f,R}$.

This definition implies that fault traces are of maximal length, i.e., a fault trace includes deviations for all residuals affected by the fault. We group the set of all fault traces into a *fault language*, represented by a *fault model* whose accepting states correspond to maximal traces.

Definition 5. (Fault Language). The *fault language* of a fault $f \in F$ with residual set R, $L_{f,R}$, is the set of all fault traces for f over the residuals in R.

Definition 6. (Fault Model). The *fault model* for a fault $f \in F$ with residual set R, is the finite automaton that accepts exactly the language $L_{f,R}$, and is given by $\mathcal{L}_{f,R} = (S, s_0, \Sigma, \delta, A)$ where S is a set of states, $s_0 \in S$ is an initial state, Σ is a set of events, $\delta : S \times \Sigma \to S$ is a transition function, and $A \subseteq S$ is a set of accepting states.

Fault signatures and RMOs can be composed automatically to form the fault models (Daigle et al., 2009). Selected fault models for a three-tank system are shown in Fig. 2. For example, as seen in $\mathcal{L}_{R_{23}^+,R}$, the fault R_{23}^+ may manifest as the fault traces $r_{p_3}^{0-} r_{p_2}^{0+} r_{p_1}^{0+}$, $r_{p_2}^{0+} r_{p_3}^{0-} r_{p_1}^{0+}$, and $r_{p_2}^{0+} r_{p_1}^{0+} r_{p_3}^{0-}$, as implied by the fault signatures and RMOs.

Fault models describe how single faults manifest in the residuals. But, candidates consist of multiple faults, so may manifest in much more complicated ways due to fault masking and the relative occurrence times of faults. The traces that result from multiple faults consist of interleavings of the fault signatures produced from the constituent faults. In our diagnosis scheme, we only observe one fault signature per residual. A second possible signature due to a different fault cannot be produced, since the model has changed since the introduction of the first fault, and therefore there is no nominal reference with which to produce the second signature. Further, the traces from multiple faults must still respect the measurement orderings of the constituent faults (Daigle et al., 2007).

As an example, take R_3^+ and R_{23}^+ with $R = \{r_{p_1}, r_{p_2}, r_{p_3}\}$ (see Fig. 2). According to the fault models, the first observed deviation must be in either r_{p_2} or r_{p_3}, as either $r_{p_3}^{0+}$ (from R_3^+), or either $r_{p_2}^{0+}$ or $r_{p_3}^{0-}$ (from R_{23}^+). Say that $r_{p_3}^{0+}$ is observed. The next deviation must then be $r_{p_2}^{0+}$ (from either R_3^+ or R_{23}^+). In the fault models, we project out the events for residuals that have already deviated, and that gives us the next set of possible events. Candidate traces continue to be built up in this way.

We can now begin to define the notion of a candidate language. We start by defining a *candidate subtrace*, which extends our earlier notion of a fault trace and is based on the notion of a *prefix*.

Definition 7. (Prefix). A trace λ_i is a *prefix* of trace λ_j, denoted by $\lambda_i \sqsubseteq \lambda_j$, if there is some (possibly empty) sequence of events λ_k that can extend λ_i s.t. $\lambda_i \lambda_k = \lambda_j$.

Definition 8. (Candidate Subtrace). Given residuals R, $\lambda = \sigma_0$ is a *candidate subtrace* for $c \subseteq F$, if $\sigma_0 \sqsubseteq \lambda' \in L_{f,R}$ for some $f \in c$. $\lambda = \lambda_i \sigma_{i+1}$ is a *candidate subtrace* for $c \subseteq F$, if λ_i is a candidate subtrace for c, and $\sigma_{i+1} \sqsubseteq \lambda' \in L_{f,R-R_i}$ for some $f \in c$, where R_i is the set of residuals that have deviated for subtrace λ_i.

We are only concerned with *maximal* traces, i.e., those for which all residuals that will deviate for the faults of the candidate have deviated (as with fault traces).

Definition 9. (Candidate Trace). Given residuals R, λ is a *candidate trace* for $c \subseteq F$ if for all $f \in c$, $L_{f,R-R_i} = \varnothing$, where R_i is the set of deviated residuals for λ.

Now, we can define the language of a candidate c, $L_{c,R}$, as the set of candidate traces for c.

Definition 10. (Candidate Language). The *candidate language* for candidate c with residual set R, $L_{c,R}$, is the set of all candidate traces for c over the residuals in R.

Similar to fault models, we can define candidate models.

Definition 11. (Candidate Model). The *candidate model* for a candidate c with residual set R, is the finite automaton that accepts exactly the language $L_{c,R}$, and is given by $\mathcal{L}_{c,R} = (S, s_0, \Sigma, \delta, A)$ where S is a set of states, $s_0 \in S$ is an initial state, Σ is a set of events, $\delta : S \times \Sigma \to S$ is a transition function, and $A \subseteq S$ is a set of accepting states.

Accepting states correspond to maximal traces. For single faults, the fault languages and fault models define the corresponding candidate languages and candidate models.

Conceptually, fault isolation works by observing the sequence of residual deviations and mapping that to consistent candidates by checking the candidate languages or by tracking the candidate models. This can be implemented efficiently online and does not require precomputation of the candidate languages or models (Daigle, 2008).

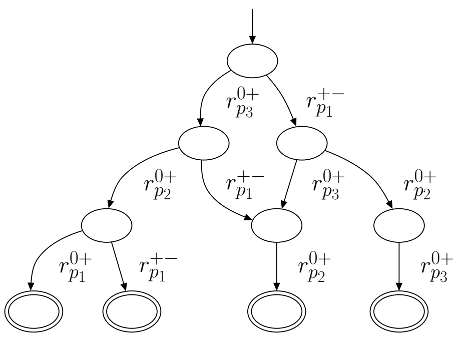


Fig. 3. Candidate model $\mathcal{L}_{C_1^- R_3^+, R}$ (obtained using the global model), where $R = \{r_{p_1}, r_{p_2}, r_{p_3}\}$.

6. DIAGNOSABILITY

Distinguishability of candidates is derived from the candidate languages. A general definition of distinguishability is as follows.

Definition 12. (Distinguishability). With residuals R, a candidate c_i is distinguishable from a candidate c_j, denoted by $c_i \nsim_R c_j$, if c_i always eventually produces effects on the residuals that c_j cannot.

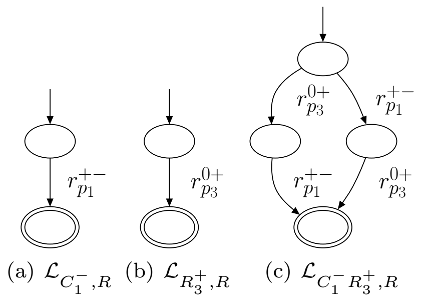
Within our framework, a basic implementation of this definition is expressed as the following proposition.

Proposition 13. (Strict Distinguishability). With residuals R, a candidate c_i is strictly distinguishable from a candidate c_j if there is no $\lambda_i \in L_{c_i,R}$ where for some $\lambda_j \in L_{c_j,R}$, $\lambda_i \sqsubseteq \lambda_j$.

For example, consider the single fault candidates shown in Fig. 2a and 2c, and the double fault candidate shown in Fig. 3, which use residuals from the global model. Clearly, C_1^- and R_3^+ are distinguishable from each other, because the first observable deviation is different for the two faults. But, C_1^- and $C_1^- R_3^+$ are not distinguishable from each other, and neither are R_3^+ and $C_1^- R_3^+$. The reason is that one fault can completely mask the other, e.g., $r_{p_1}^{+-} r_{p_2}^{0+} r_{p_3}^{0+}$ may be observed either because C_1^- has occurred by itself, or because C_1^- and R_3^+ have both occurred, and R_3^+ has been completely masked.

But, the decoupling introduced by PCs can eliminate some of this masking. Fig. 4 shows the candidate models for these candidates with the PC-based residuals. Since r_{p_1} is decoupled from R_3^+, and r_{p_3} is decoupled from C_1^-, if both faults occur together, we see deviations in both residuals and either fault by itself will not be consistent. Therefore, $C_1^- R_3^+$ is distinguishable from both C_1^- and R_3^+. However, the converse is still not true, i.e., C_1^- is not distinguishable from $C_1^- R_3^+$. If C_1^- occurs, then we see $r_{p_1}^{+-}$, which so far, is consistent with both the single and the double fault. We then have to wait infinitely long to ensure that r_{p_3} does not deviate and confirm that C_1^- has occurred by itself, and so we say they are not distinguishable.

In practice, however, this is a fairly strong distinguishability requirement to be working with. If C_1^- occurs and



(a) $\mathcal{L}_{C_1^-,R}$ (b) $\mathcal{L}_{R_3^+,R}$ (c) $\mathcal{L}_{C_1^- R_3^+,R}$

Fig. 4. Candidate models (obtained using the PCs) where $R = \{r_{p_1}, r_{p_2}, r_{p_3}\}$.

we see $r_{p_1}^{+-}$, then, since it is not possible for the effects of C_1^- to mask the effects that will be produced by R_3^+, if it happens to occur, we will definitely see evidence for it (i.e., $r_{p_3}^{0+}$). So, before we see such evidence, we are safe in assuming that R_3^+ has not also occurred. In this case we want to be able to say that C_1^- is distinguishable from $C_1^- R_3^+$. This leads to a new implementation of distinguishability.

Proposition 14. (Partial Masking Distinguishability).With residuals R, a candidate c_i is distinguishable under partial masking from a candidate c_j if (i) $c_i \subset c_j$ and $R_{c_i} \cap R_{c_j - c_i} = \varnothing$ or if (ii) c_i and c_j are strictly distinguishable.

Here, the notation $R_c \subseteq R$ refers to the subset of the residuals that will deviate due to candidate c. Applied to our example in Fig. 4, this says that, since it is not possible for C_1^- to mask any of the residuals that R_3^+ affects, we can say C_1^- is distinguishable from $C_1^- R_3^+$ (since $C_1^- \subset C_1^- R_3^+$ and $R_{C_1^-} \cap R_{C_1^- R_3^+ - C_1^-} = \varnothing$). Similarly, we can say R_3^+ is distinguishable from $C_1^- R_3^+$. A weaker version of this distinguishability implementation allows partial masking but disallows complete masking.

Proposition 15. (Complete Masking Distinguishability). With residuals R, a candidate c_i is distinguishable under complete masking from a candidate c_j if (i) $c_i \subset c_j$ and $R_{c_j - c_i} \nsubseteq R_{c_i}$ or if (ii) c_i and c_j are strictly distinguishable.

For example, if both C_1^- and R_3^+ also affected some other residual other than r_{p_1} and r_{p_3}, and C_1^- occurs and we see $r_{p_1}^{+-}$, then even if we see this other residual deviate, we are somewhat safe in assuming R_3^+ has not yet occurred, because if it does we will eventually see evidence for it.

We can improve distinguishability even more if we can confidently observe the lack of a deviation, e.g., by assuming that faults will cause residual deviations at most x seconds after they occur. That is, if a fault should affect some residuals r_1 and r_2 and we have observed r_1 deviate, but r_2 has not deviated x seconds since r_1 was observed to deviate, then we can assume that fault has not occurred[3]. In this case, distinguishability checks only for common candidate traces and the prefix of traces does not matter.

Proposition 16. (Weak Distinguishability). With residuals R, a candidate c_i is weakly distinguishable from a candidate c_j if $L_{c_i,R} \cap L_{c_j,R} = \varnothing$.

Table 3. Diagnosability results for $F = \{C_1^-, C_2^-, C_3^-, R_1^+, R_2^+, R_3^+, R_{12}^+, R_{23}^+\}$.

M	Distinguishability	Global model	PCs	Combined
M_p	Strict	522 (0.41)	196 (0.16)	179 (0.14)
	Partial masking	522 (0.41)	164 (0.13)	147 (0.12)
	Complete masking	522 (0.41)	154 (0.12)	137 (0.11)
	Weak	522 (0.41)	62 (0.05)	62 (0.05)
M_q	Strict	448 (0.36)	335 (0.27)	305 (0.24)
	Partial masking	448 (0.36)	329 (0.26)	299 (0.24)
	Complete masking	448 (0.36)	314 (0.25)	284 (0.23)
	Weak	448 (0.36)	314 (0.25)	284 (0.23)

With distinguishability defined, we can now begin to define diagnosability. Diagnosability assumes a given implementation of distinguishability. It depends on the set of candidates being considered and the chosen set of residuals. First we define a *system.*

Definition 17. (System). A *system* $\mathcal{S}$ is a tuple (F, C, M, R, $L_{C,R}$), where F is a set of faults, $C = \{c_1, c_2, \ldots, c_n\} \subseteq 2^F$ is a set of candidates, M is a set of measurements, R is a set of residuals, and $L_{C,R} = \{L_{c_1,R}, L_{c_2,R}, \ldots, L_{c_n,R}\}$ is the set of candidate languages.

Here, the set of candidates does not have to be the full powerset 2^F, e.g., it may include only single faults, single faults and double faults, etc.

A system is diagnosable if all pairs of candidates are distinguishable for the given implementation of distinguishability. If diagnosable, then we can make guarantees about the unique isolation of every candidate in the system.

Definition 18. (Multiple Fault Diagnosability). A system $\mathcal{S} = (C, F, M, R, L_{C,R})$ is *diagnosable* if $(\forall c_i, c_j \in C)$ $c_i \neq c_j \implies c_i \not\sim_R c_j$.

Even with PCs, in many cases we do not expect complete diagnosability, therefore, we introduce a diagnosability score in order to compare different approaches. For a candidate set C, the score is computed as the number of indistinguishable candidate pairs. The worst possible score is $2\binom{|C|}{2}$.[4] We compute the normalized score, describing the fraction of undiagnosability, as the diagnosability score divided by the worst score.

7. RESULTS

As a first scenario, consider the three-tank system with $F = \{C_1^-, C_2^-, C_3^-, R_1^+, R_2^+, R_3^+, R_{12}^+, R_{23}^+\}$ and two different measurement sets $M_p = \{p_1, p_2, p_3\}$ and $M_q = \{q_1, q_2, q_3\}$. Table 3 shows the diagnosability results using both measurement sets for PCs and the global system model for each one of the four distinguishability definitions. For this example, the worst possible score is $2\binom{36}{2} = 1260$. In the table, the columns show the measurement set used, the distinguishability definition, and scores for the global model and PC approaches, respectively. Normalized scores are shown in parentheses.

From the results, diagnosability using PCs is clearly much better than using the global model. Also, the improvement is much more substantial for M_p than for M_q, since measurement set M_p provides more decoupling. This is

[3] In practice, this can be difficult to achieve because it would be affected by fault magnitude, sensor noise, and properties of the fault detectors.

[4] The factor of 2 appears because distinguishability is not a symmetric property.

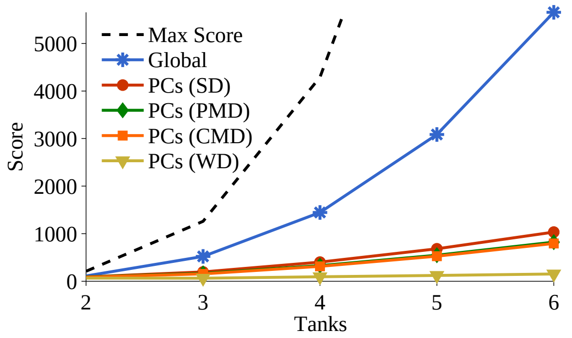


Fig. 5. Scalability of diagnosability.

consistent with the intuition that decoupling improves diagnosability. Comparing the distinguishability definitions, we see that using weak distinguishability is the best in all cases, followed by complete masking, partial masking, and strict distinguishability. This is expected, since weak distinguishability definition is least restrictive, and then complete masking, partial masking, and strict.

Here, many candidates can be distinguished using PCs compared to the global model. However, there is also a small subset of candidates that can be distinguished using the global model, but not using the PCs. For example, consider distinguishing $C_1^- R_1^+$ from R_{12}^+. With the global model, we can distinguish these candidates because they produce different effects on r_{p_2} and r_{p_3} (see Table 1). With PCs, however, if $C_1^- R_1^+$ occurs and we see $r_{p_1}^{0+}$, then R_{12}^+ remains consistent and we will not see another deviation in order to eliminate it, so they are not distinguishable with strict, partial masking, and complete masking definitions.

Therefore, improvements in diagnosability can be achieved in an approach that combines the residual sets from both the global model and the PCs. In such an approach, two candidates are distinguishable if they are distinguishable using either the global model-based residuals or the PC-based residuals. The fifth column of Table 3 provides the diagnosability results in this case, confirming that in all the cases, a combined approach provides results equal to or better than the approach with only PC-based residuals.

Although we cannot obtain complete multiple fault diagnosability in this case, sometimes it can be achieved. For example, consider $F = \{C_1^-, C_2^-, C_3^-\}$ and $M_p = \{p_1, p_2, p_3\}$. Here, it is not diagnosable for PCs with strict distinguishability (score of 6 (0.2)) and the global model (score of 18 (0.6)), but otherwise we get perfect diagnosability. This occurs because the faults are completely decoupled from each other by the PCs. In fact, whenever the measurement set is such that we get full decoupling with PCs, we will always achieve perfect diagnosability for any of the distinguishability definitions.

It is also interesting to investigate the scalability of these diagnosability properties. We computed diagnosability scores for 2-6 tanks (see Fig. 5). The worst possible score increases significantly as the number of tanks increases, because each tank adds three new faults to the system. The scores for the global approach increase as well, but at a significantly smaller rate. For the PC-based diagnosability results, the growth rate is reduced even further. In fact, when using weak distinguishability, the scores for the PC-based approach, for 3 tanks and higher, increase at a linear rate, with only 30 new indistinguishable candidate pairs being added for each new tank. Clearly, diagnosability scales much better with the PC-based approach.

8. CONCLUSIONS

In this work, we have presented a qualitative, event-based framework for multiple fault isolation with PCs. The decoupling of faults from residuals provided by PCs leads to a great improvement in multiple fault diagnosability since the possibility of fault masking, when multiple faults occur, is reduced. We have established a definition for multiple fault diagnosability within our framework, providing several notions of distinguishability. Diagnosability analysis of a system may then be used to determine the expected amount of ambiguity after QFI, and which ambiguities will need to be resolved by more expensive quantitative methods.

Experimental results on a multi-tank system show the improvement of multiple fault diagnosability when PCs are used instead of the global system model. Moreover, using a combined approach of global model- and PC-based residuals, we obtain further improvements in diagnosability. Diagnosability is also more scalable with the PC-based approach, and in fact, diagnosability scores grow only linearly for the tank system using weak distinguishability.

In this paper, we considered only single and double faults for the case study, but, in future work, we will study how the approach scales with candidates of higher cardinality. Also, we will extend this approach to develop an MFD framework including multiple fault identification.

REFERENCES

Daigle, M. (2008). *A Qualitative Event-based Approach to Fault Diagnosis of Hybrid Systems*. Ph.D. thesis, Vanderbilt University.

Daigle, M., Koutsoukos, X., and Biswas, G. (2007). A qualitative approach to multiple fault isolation in continuous systems. In *Proc. of the Twenty-Second AAAI Conference on Artificial Intelligence*, 293–298.

Daigle, M.J., Koutsoukos, X., and Biswas, G. (2009). A qualitative event-based approach to continuous systems diagnosis. *IEEE Trans. Ctrl. Sys. Tech.*, 17(4), 780–793.

de Kleer, J. and Williams, B.C. (1987). Diagnosing multiple faults. *Artificial Intelligence*, 32, 97–130.

Dvorak, D. and Kuipers, B. (1991). Process monitoring and diagnosis: a model-based approach. *IEEE Expert*, 6(3), 67 –74.

Gertler, J. (1998). *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, New York.

Mosterman, P.J. and Biswas, G. (1999). Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans. Syst. Man Cy. A.*, 29(6), 554–565.

Nyberg, M. and Krysander, M. (2003). Combining AI, FDI, and statistical hypothesis-testing in a framework for diagnosis. In *Proc. of IFAC Safeprocess'03*.

Pulido, B. and Alonso-González, C. (2004). Possible Conflicts: a compilation technique for consistency-based diagnosis. *IEEE Trans. on Systems, Man, and Cybernetics, Part B*, 34(5), 2192–2206.