A Consistency-based Robust Diagnosis Approach for Temporal Causal Systems*

Sherif Abdelwahed **Gabor Karsai** **Gautam Biswas**
Institute for Software Integrated Systems,
Vanderbilt University, Nashville, TN, 37203
{sherif,gabor,biswas}@isis.vanderbilt.edu

Abstract

In this paper we present a consistency-based robust diagnosis approach for a class of temporal causal systems modeled as timed failure propagation graphs. Timed failure propagation graphs are causal models that capture the temporal characteristics of failure propagation in dynamic systems. In this paper, we define the problem of robust diagnosis for this class of systems and introduce an optimal diagnosis algorithm that is robust with respect to sensor faults. The paper outlines the proof for the correctness and optimality of the proposed algorithm.

1 Introduction

Diagnostic algorithms detect, isolate, and estimate system failures using observed signals and measurements from the system sensors and actuators. Comparing these against a model that captures nominal and/or faulty behavior produces fault hypotheses that explain the observed system condition.

In many industrial systems, diagnosis is limited to signal monitoring and fault identification via threshold logic, e.g., detecting if a sensor value deviates from its nominal value. Failure propagation is modeled by capturing the qualitative association between sensor signals in the system for a number of different fault scenarios. Typically, such associations correspond to relations used by human experts in detecting and isolating faults. This approach has been effectively used for many complex engineering systems. Common industrial diagnosis methods include fault trees [Himmelblau, 1978; Viswanadham and Johnson, 1988; Hessian *et al.*, 1990; Ishida *et al.*, 1985], cause-consequence diagrams [Rao and Viswanadham, 1987a; 1987b], diagnosis dictionaries [Richman and Bowden, 1985], and expert systems [Scherer and White, 1989; Tzafestas and Watanabe, 1990].

Model-based diagnosis (see [Frank, 1990; Hamscher *et al.*, 1992; Patton, 1994] and the references therein), on the other hand, compares observations from the real system with the predictions from a model. Analytical models, such as state equations, finite state machines, and predicate/temporal logic are used to describe the nominal system behavior. In the case of a fault, discrepancies between the observed behavior and the predicted normal behavior occur. These discrepancies can then be used to detect, isolate, and identify the fault depending on the type of model and methods used.

This paper presents a consistency-based approach for robust diagnosis of dynamic systems in which failure behavior can be captured by a class of temporal causal models. Consistency based diagnosis was introduced in a logical framework in [Reiter, 1987] and was later extended in [de Kleer *et al.*, 1992]. In consistency-based diagnosis the behavior of the system is predicted using a nominal system model and then compared with observations of the actual behavior of the system to obtain the minimal set of faulty component that is consistent with the observations and the nominal model. Consistency based diagnosis have been applied to develop diagnosis algorithms for causal systems [Darwiche, 1998; Darwiche and Provan, 1996] and temporal causal systems [Gamper, 1996; Console and Torasso, 1991].

The diagnosis approach presented in this paper is conceptually related to the temporal causal network approach presented in [Console and Torasso, 1991]. However, we focus on incremental reasoning and diagnosis robustness with respect to sensor failures. The causal model presented in this paper is based on the timed failure propagation graph (TFPG) introduced in [Misra, 1994; Misra *et al.*, 1994]. The TFPG model is closely related to fault models presented in [Padalkar *et al.*, 1991; Karsai *et al.*, 1992; Mosterman and Biswas, 1999] and used for an integrated fault diagnoses and process control system [Karsai *et al.*, 2003]. The TFPG model was extended in [Abdelwahed *et al.*, 2004] to include mode dependency constraints on the propagation links, which can then be used to handle failure scenarios in hybrid and switching systems. The extended model is referred to as a Hybrid Failure Propagation Graph (HFPG).

In this paper, we introduce the main elements of the robust diagnosis problem for a simplified version of the HFPG model, referred to as Simple Timed Failure Propagation Graphs (sTFPG), with a disjunctive propagation dependency, all monitored discrepancies, and no mode switching. The proposed algorithm is robust (degrades gracefully) with respect to sensor failures. We formally describe an incremental optimal diagnosis procedure for this class of models. The proposed algorithm consists of two main procedures. The

*Funded, in part, by Boeing and the NASA ALS program (Contract: NCC 9-159).

first one generates consistent hypothesis from an initial state assignment while the second generates an optimal consistent initial state assignment based on current state observation.

The paper is organized as follows. In Section 2, the simple timed failure propagation graph model is introduced. Section 3 introduces the main elements of the diagnosis problem for timed causal systems modeled as sTFPG. In this section, optimal diagnosis for sTFPG is defined based on of the notion of observed and hypothetical states and the matching between them. Section 4 presents the optimal diagnosis algorithm.

2 Simple timed failure propagation graphs

A sTFPG is a labeled directed graph where the nodes represent either failure modes, which are fault causes, or discrepancies, which are off-nominal conditions that are the effects of failure modes. Edges between nodes in the graph capture propagation of failure effects over time in the dynamic system. Formally, a sTFPG model is represented as a tuple $G = (F, D, E, \mathrm{tmin}, \mathrm{tmax})$, where:

- F is a set of failure nodes
- D is a set of discrepancy nodes, with $F \cap D = \varnothing$
- $E \subset V \times V$ is a set of edges, where $V = F \cup D$
- $\mathrm{tmin}, \mathrm{tmax} : E \to \mathbb{R}$ assign to each edge $e \in E$ its minimum and maximum propagation time, respectively.

An edge $e = (v, v') \in E$ indicates that a state change can occur from v to v' due to propagation effects. For an edge $e \in E$, we will use the notation $e.\mathrm{tmin}$ and $e.\mathrm{tmax}$ to indicate the corresponding minimum and maximum time for failure propagation along the edge e, respectively. This implies that for $(v, v') \in E$ it will take at least (at most) $(v.v').\mathrm{tmin}$ ($(v.v').\mathrm{tmax}$) time for the fault to propagate from v to v'. For all $e \in E$ we assume that $0 \leq e.\mathrm{tmin} \leq e.\mathrm{tmax}$.

We assume that sTFPG models do not contain self loops and that failure modes are root nodes, i.e., they cannot be a destination of any edge. Also, every discrepancy must be a successor of another discrepancy or a failure mode.

Figure 1 shows example of an sTFPG model. In this figure rectangles represent the failure modes while bold-line circles represent discrepancies. Edges between nodes capture failure/discrepancy propagation in the system. Propagation through edges are parameterized by the interval $[e.\mathrm{tmin}, e.\mathrm{tmax}]$. Sequences of alarms are identified by shaded discrepancies. The time at which the alarm is observed is shown above the corresponding discrepancy.

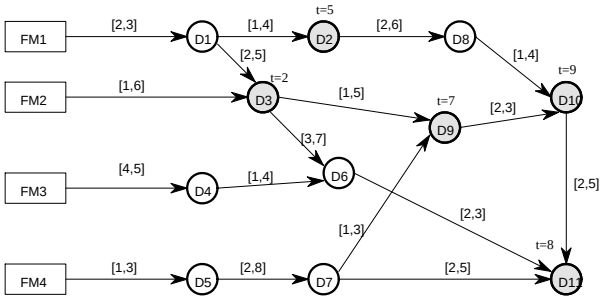

Figure 1: A simple timed failure propagation graph

The sTFPG model captures observable failure propagations between sensors in practical systems. In this setting, sensors capture state deviations from nominal values. The set of all observed deviations corresponds to the discrepancy set in the sTFPG model. The propagation edges corresponds to causality (for instance, corresponding to energy flow) in the system dynamics. Due to the dynamic nature of the system, failure effects take time to propagate between the system components (such time in general depend on the system's time constants as well as the size and time evolution of underlying failure). In many practical situations such delays can be computed analytically or by simulation of an accurate model.

Failure propagation in a sTFPG has a simple semantics. The state of a node indicates if the failure effects reached this node. Failure effects can reach a node from any of its predecessors. Assuming $e = (v, v') \in E$, then once a failure effect reaches v at time t it must reaches v' at a time t' where $e.\mathrm{tmin} \leq t' - t \leq e.\mathrm{tmax}$. Once a failure effect reaches v' from any of its predecessors its state will change permanently, and it will not be affected by any future failure propagation.

In the rest of this paper we consider the scenario presented in Figure 1. This scenario corresponding to failure FM2 occurring at time $t = 1$, and two sensor failures; D2 (false alarm) and D6 (missing alarm).

3 The Diagnosis Problem

An *actual (physical) system* state corresponds to the current state of all nodes in the sTFPG model. Formally, a physical state a time t is a map $AS_t : V \to \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R}$, where V is the set of nodes in the sTFPG model. Therefore, $AS(v)$ denotes the state of a node v. Here a state ON indicates that the failure effect reached this node, otherwise it is set to OFF. We will write $AS(v).\mathrm{state}$ to indicate the first element of the function (the state) and $AS(v).\mathrm{time}$ to indicate the second element (state change time). For instance in the scenario described for Figure 1, we have $AS_5(\mathrm{FM2}) = \{\texttt{ON}, 1\}$. Given that failure effects are permanent, the state of a node once changed will remain constant after that.

The aim of the diagnosis process is to identify the current actual state of the system. In the sTFPG setting, however, discrepancies are observable while failure modes are not. In addition, due to possible sensor failures, the observed state may not be consistent with the sTFPG model constraints. Therefore, the actual state cannot be identified with absolute certainty. The diagnosis process will then try to find an estimate of the current state of the system that is consistent with the observations, and is as close as possible to the observed state. In this section, the diagnosis problem is formally defined based on the notions of observed and hypothetical states, state consistency, and observation matching.

3.1 Observed states

An *observed state* at time t is defined as a map $S_t : D \to \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R}$. Observed states are only defined for discrepancies. Similar to Actual states, the map S_t is unique for each time instance t. We assume that sensor signals are permanent so that the observed state of a discrepancy once changed will remain constant after that. This also applies to faulty sensors.

Observed states, due to potential sensor failures, may not be consistent with the failure propagation graph model temporal constraints. Here, consistency is defined in terms of the causality and propagation timing information expressed in the sTFPG model. In particular, observed state consistency is defined as a binary relation on the set of observed states for adjacent nodes at a given time. Formally, for two nodes $d', d \in D$ we say that d is timing consistent, or simply *T-consistent*, with d' at time t if $(d', d) \in E$ and any of the following holds:

1. $S_t(d').\text{state} = \texttt{OFF}$ and $S_t(d).\text{state} = \texttt{OFF}$,
2. $S_t(d').\text{state} = \texttt{ON}$ and $S_t(d).\text{state} = \texttt{OFF}$ and $t < S_t(d').\text{time} + (d', d).\text{tmax}$,
3. $S_t(d').\text{state} = \texttt{ON}$ and $S_t(d).\text{state} = \texttt{ON}$ and $(d', d).\text{tmax} \geq S_t(d).\text{time} - S_t(d').\text{time} \geq (d', d).\text{tmin}$

The above conditions simply state that d', d are T-consistent at time t if their observed states do not contradict the propagation temporal constraints defined by the sTFPG model. The T-consistency relation at time t is denoted C_t. For example, in Figure 1, $C_0 = C_2 = D \times D$. At time $t = 9$, we have $\{(D4, D6), (D5, D7), (D3, D9), (D9, D10)\} \subset C_9$ while $(D3, D6) \notin C_9$, although $(D3, D6) \in C_8$. We also need to define the notion of weak consistency. For two nodes $d', d \in D$ where we say that d is weakly T-consistent, or simply *WT-consistent*, with d' at time t if $(d', d) \in E$ and any of the following holds:

1. $S_t(d').\text{state} = \texttt{OFF}$ and $S_t(d).\text{state} = \texttt{ON}$,
2. $S_t(d').\text{state} = \texttt{ON}$ and $S_t(d).\text{state} = \texttt{ON}$ and $S_t(d).\text{time} < (d', d).\text{tmin} + S_t(d').\text{time}$

The WT-consistency relation at time t is denoted W_t. For example, $(D1, D2) \in W_5$. Note that $C_t \cap W_t = \varnothing$ and in general C_t, W_t are not a cover for $D \times D$. Now we extend the notion of consistency to arbitrary discrepancy sets. A set of discrepancies $D' \subseteq D$ is said to be T-consistent if for all $d', d \in D'$ with $(d', d) \in E$:

- $(d', d) \in C_t$, or
- $(d', d) \in W_t \wedge ((\exists d'' \in D')(d'', d) \in E \wedge (d'', d) \in C_t)$

Note that the above condition is true by default for any non-adjacent discrepancies. We define the set of *supports* at time t, to be the set all T-consistent sets with maximum size (number of discrepancies). This set will be denoted Ψ_t and the size of its sets will be denoted λ_t. For example, in Figure 1, $\Psi_0 = \Psi_1 = D$, $\Psi_4 = \Psi_7 = \{D - \{D1\}, D - \{D2\}\}$. Note that the set of all T-consistent set is a cover for D, and that $\Psi_0 = D$ in any sTFPG.

In temporal causal systems, sensor signals are received sequentially. In this case, the set Ψ_t can be computed efficiently by maintaining a list of T-consistent sets, and updating this list when a new sensor signal is received or a time-out event is generated automatically indicating a missing sensor signal (according to the current list of T-consistent sets).

The set of T-consistent sets can be iteratively updated as follows. Let d be the discrepancy corresponding to a new sensor signal, and let $TC(d)$ be the current set of all T-consistent sets that contains d. Then for any $D' \in TC(d)$, if the new observed state of d is consistent with D then d remains in D'. Otherwise, d must be inconsistent with one or more of its adjacent nodes. In this case the set D is split, such that the conflicting nodes are separated. New T-consistent sets are generated by removing conflicting nodes.

3.2 Hypothetical states

An *hypothetical state* defines node states and the interval at which each node changes its state. In contrast to the observed state, a hypothetical state is defined for all the sTFPG nodes. Formally a hypothetical state is a map $H_t : V \rightarrow \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R} \times \mathbb{R}$. Similar to actual states, hypothetical states are defined for both discrepancies and failure modes. We will write $H(v).\text{terl}$ to indicate the estimated earliest time of state change and $H(v).\text{tlat}$ to indicate the estimated latest time. Similar to actual and observed states, a node v with $\texttt{OFF}$ state has activation time of exactly 0, that is $H(v).\text{state} = \texttt{OFF}$ implies $H(v).\text{terl} = H(v).\text{tlat} = 0$.

Hypothetical states are an estimation of the current state of all nodes in the system and the time period at which this node changed its states. An estimation of the current state, however, must be consistent with the sTFPG model to be useful. Similar to observed states, the consistency of hypothetical states depends on the underlying sTFPG model and the current time. Formally, we say the hypothetical state H_t is *consistent* if for any node d the following conditions hold:

1. $H_t(d) = \texttt{OFF}$ and for all $(v, d) \in E$:
   (a) $H_t(v) = \texttt{OFF}$, or
   (b) $H_t(v) = \texttt{ON} \wedge t < H_t(v).\text{tlat} + (v, d).\text{tmax}$
2. $H_t(d) = \texttt{ON}$ and all the following hold:
   (a) $H_t(d).\text{terl} \geq \min_{v \in U_d}\{H_t(v).\text{terl} + (v, d).\text{tmin}\}$,
   (b) $H_t(d).\text{tlat} \leq \min_{v \in U_d}\{H_t(v).\text{tlat} + (v, d).\text{tmax}\}$

   $U_d = \{v \in V | (v, d) \in E \text{ and } H_t(v).\text{state} = \texttt{ON}\}$

The above simply says that H is consistent if every node state is consistent with the state of its predecessors. Namely, the state of a node d can be $\texttt{OFF}$ at time t if there is a possibility that a failure did not propagate and reach d at time t from any of its predecessors. Also, the state of a node d can be $\texttt{ON}$ at time t if there is a possibility that a failure have reached d at or before t from any of its predecessors. A consistent hypothetical state will be referred to as a *hypothesis*. In general, there are infinitely many consistent hypothesis at any time t.

A consistent hypothetical state is generated from an initial state assignment to a subset of the system nodes. Formally, let $V' \subset V$ be a set of nodes. A partial hypothetical state assignment at time t is a map $PH_t : V' \rightarrow \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R} \times \mathbb{R}$. The map PH_t is said to be consistent for V' if the PH is consistent with respect to the subgraph of the sTFPG model restricted to the V' nodes. In this case it is referred to as a partial hypothesis. A hypothesis defined over V may be referred to as a complete hypothesis.

For an initial state assignment PH_t, an *extension* to PH_t is a (complete) hypothesis H_t satisfying $H_t|_{V'} = PH_t$, where $H_t|_{V'}$ is the restriction of the map H_t to the subset V'. We say that PH_t is a *valid state assignment* if there exists at least one extension for PH_t. Given that PH_t is a valid state assignment for $V' \subseteq V$ one can construct an extension H by

assigning state values to the remaining nodes, $V-V'$ by propagating the current state assignment backward and forward to adjacent states. The procedure for hypothesis generation will be explained later in more detail.

We define a state equivalence relation of hypothesis as follows. Let H and H' be two hypothetical states at time t. We say that H and H' are state equivalent if

$$(\forall v \in V) \quad H(v).\text{state} = H'(v).\text{state}$$

The state equivalence relationship is denoted $\equiv_s$. It is easy to see that the equivalence kernel of this relationship is finite. Namely, the set of all state-equivalent hypothesis sets is finite In fact, the equivalence kernel of $\equiv_s$ is isomorphic to the set of all binary assignments to V.

We now define an order relation on state equivalent hypotheses (equivalence classes of $\equiv_s$). Let H and H' be two hypothetical states at time t. Then we say that H' is contained in H, written as $H' \prec H$ if:

$$(\forall v \in V) \ [H'(v).\text{terl}, H'(v).\text{tlat}] \subset [H(v).\text{terl}, H(v).\text{tlat}]$$

We extend this relation to the partial order $\preceq$ in the usual way (by including the identity relation restricted to the underlying equivalence class). Clearly, $H' \preceq H$ implies $H' \equiv_s H$. A hypothesis H is said to *maximal* if there is no other hypothesis H' such that $H \prec H'$. In other words, any hypothetical state H' satisfying $H \prec H'$ is not consistent.

Proposition 1 For any hypothesis H there exists a unique maximal hypothesis H^* such that $H \preceq H^*$. □

The proof of the above proposition is direct based on the definition of consistent hypothetical states. The maximal hypothesis containing H can be computed by iteratively extending the state change period for each node while maintaining consistency between the states.

Let PH_t be a valid state assignment, and let H_t be a hypothesis. We say that H_t is maximal for PH_t if H_t is an extension for PH_t and for any other extension H'_t that is state-equivalent to H_t we have $H'_t \prec H_t$.

3.3 The diagnosis set

As mentioned earlier, the task of the diagnosis process is to provide a set of consistent state estimation (hypotheses) that closely matches the observed state of the system. We say that a hypothesis H_t strongly matches a discrepancy d if

- $H_t(d).\text{state} = S_t(d).\text{state}$, and
- $H_t(d).\text{terl} = H_t(d).\text{tlat} = S_t(d).\text{time}$

The set of all discrepancies that are strongly matched with a hypothesis H is denoted $SM(H_t)$. A hypothesis H_t weakly matches a discrepancy d if:

- $H_t(d).\text{state} = S_t(d).\text{state}$, and
- $H_t(d).\text{terl} \leq S_t(d).\text{time} \leq H_t(d).\text{tlat}$

or

- $H_t(d).\text{state} = \texttt{ON} \ \wedge \ S_t(d).\text{state} = \texttt{OFF}$, and
- $t \leq H_t(d).\text{tlat}$

or

- $H_t(d).\text{state} = \texttt{OFF} \ \wedge \ S_t(d).\text{state} = \texttt{ON}$, and
- $H_t(d).\text{terl} \leq t < H_t(d).\text{tlat}$

The set of all discrepancies that are weakly matched with a hypothesis H is denoted $WM(H_t)$. Clearly, for any hypothesis H_t we have $SM(H_t) \subseteq WM(H_t)$. The rank, $R(H_t)$ of hypothesis H_t is defined as the size of its weakly matched set $WM(H_t)$. Clearly, for a sTFPG with N discrepancies we have $R(H_t) \leq N$.

A maximal hypothesis with the highest rank at time t is referred as a *diagnosis*. We will write R_t to denote the highest rank at time t. The *diagnosis set* at time t as the set of maximal hypothesis with rank R_t. The *diagnosis problem* can then be stated as follows:

> Given a set of sensor signals at time t, compute the diagnosis set (hypotheses with maximal ranks) corresponding to the underlying observed state.

Assuming a set of uncorrelated sensors that are more likely not to fail (failure rate is less than 50%), it is easy to see that the diagnosis set defined above is the most likely estimation of the actual state of the system nodes based on the observed state.

4 The Diagnosis Reasoning Algorithm

The aim of the diagnosis reasoning process is to find a consistent and plausible explanation (in the form of a hypothetical state) of the current system state based on sensor measurements (observed state). Typically, the system starts with no sensor signals (all discrepancies are in the OFF state). When a failure occur, sensor signals are generated and received sequentially by the diagnosis reasoner, which in turn reacts by generating the corresponding hypotheses, from which the state of failure modes can be identified.

4.1 Hypothesis generation

A hypothesis must be generated from an initial state assignment for at least one of the system nodes. As discussed earlier, such state assignments must be valid. By definition, a valid initial state assignment is consistent within its domain of nodes and can be extended to a hypothesis. However, apart from the consistency assumption, such definition of validity is not constructive, namely, one cannot test or generate a valid state assignment from this definition.

One way to ensure validity of a consistent hypothetical state assignment $PH : V' \rightarrow \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R} \times \mathbb{R}$ is to restrict the elements of the set V' such that they do not have common ancestors from $V - V'$, that is, they are causally independent with respect to the remaining nodes. Formally, a subset V' of V is *causally independent* if for all $v \in V'$:

1. $(\exists v' \in V') \quad (v', v) \in E$, or
2. $(\forall v' \in V')(\forall v'' \in V - V') \quad \neg(\{v, v'\} \subseteq \text{Reach}(v''))$

where $\text{Reach}(v'')$ is the set of all nodes reachable from v'' in the sTFPG graph. Intuitively, V' is causally independent if the state of each node in V' can be either explained by the state of another node in V' or does not share a common ancestor from outside V' with another node in V'.

Let PH be a consistent state assignment for a causally independent set $V' \subseteq V$. We define the backward extension operation, BProp, on the set V', the assignment map PH, and a node $v \in V - V'$ where $(v, v') \in E$ for some $v' \in V'$ such that v' does not have a predecessor in V'. The outcome of this operation is a new state assignment map PH' for the set $V' \cup \{v\}$, where $PH'|_{V'} = PH$ and $PH'(v)$ is assigned as follows:

- If $PH(v').\text{state} = \texttt{OFF}$:
  - $PH'(v).\text{state} = \texttt{OFF}$, $PH'(v).\text{terl} = PH'(v).\text{tlat} = 0$
- If $PH(v').\text{state} = \texttt{ON}$:
  - $PH'(v).\text{state} = \texttt{ON}$,
  - $PH'(v).\text{terl} = PH(v').\text{terl} - (v, v').\text{tmax}$,
  - $PH'(v).\text{tlat} = PH(v').\text{tlat} - (v, v').\text{tmin}$

It is easy to verify that the state assignment PH' is consistent for $V' \cup \{v\}$ given that PH is consistent for V'. In addition we have the following result.

Proposition 2 In the BProp operation described above, the set $V' \cup \{v\}$ is causally independent and the state assignment PH' is maximal with respect to PH.

Proof (Outline): Assume that $V'' = V' \cup \{v\}$ is not causally independent, then there must exist two nodes d, d' that do not have a predecessor in V'' and share a common ancestor from $V - V''$. Clearly one of these nodes must be v. However, if v shares a common ancestor with a node $d \in V'$ then v' must also share the same ancestor with d contradicting the assumption that V' is causally independent. The proof of maximality is direct from the definition of consistent hypothesis. □

A direct consequence of the above result is that, iterative applications of the BProp operation on a consistent hypothetical assignment PH over a causally independent set will result in a maximal consistent hypothetical state assignment for PH, over a larger causally independent set. Also, clearly that iterative application of BProp will terminate at certain point at which each node in the underlying set has either a predecessor from inside the set or no predecessor at all in the sTPFG model (a failure mode in this case).

Let V' be a causally independent set of nodes and PH be a consistent state assignment for V'. We define the forward extension operation, FProp, on the set V', the assignment map PH, and a node $v \in V - V'$ where $(v', v) \in E$ for some $v' \in V'$ and in addition:

$$(\forall v'' \in V - V')\ (v'', v) \in E \rightarrow (\forall d \in V')v'' \notin \textsf{Reach}(d)$$

From the above, the external node v should have a predecessor in V'. In addition, any other predecessor of v that is not in V' should not be reachable from a node in V'. These conditions on v are required to ensure correct causal propagation. The outcome of the FProp operation is a new state assignment map PH' for the set $V' \cup \{v\}$, where $PH'|_{V'} = PH$ and and $PH'(v)$ is assigned as follows:

- If $(\forall v' \in V')\ (v', v) \in E \rightarrow PH(v').\text{state} = \texttt{OFF}$:
  - $PH'(v).\text{state} = \texttt{OFF}$, $PH'(v).\text{terl} = PH'(v).\text{tlat} = 0$
- Else:
  - $PH'(v).\text{state} = \texttt{ON}$,
  - $PH'(v).\text{terl} = \min_{v' \in U_v}\{PH(v').\text{terl} + (v, v').\text{tmin}\}$,
  - $PH'(v).\text{tlat} = \min_{v' \in U_v}\{PH(v').\text{tlat} + (v, v').\text{tmax}\}$

  $U_v = \{v' \in V' | (v', v) \in E \text{ and } PH(v').\text{state} = \texttt{ON}\}$

It is easy to verify that the state assignment PH' is consistent for $V' \cup \{v\}$ given that PH is consistent for V'. Similar to the case of backward propagation we have the following result.

Proposition 3 In the FProp operation described above, the set $V' \cup \{v\}$ is causally independent and the state assignment PH' is maximal with respect to PH.

Proof (Outline): V'' only adds a node that has a predecessor in V' so it inherits the consistency of V'. The maximality of PH' is direct from the definition of consistency. □

Similar to the case of backward propagation, the above result shows that iterative applications of the FProp operation on a consistent hypothetical assignment PH over a causally independent set will result in a maximal consistent hypothetical state assignment for PH, over a causally independent set. Also, it is easy to see that the iterative application of FProp will terminate after finite number of steps at which all the decedents of V' are assigned a state value.

Based on the above operations we can now define a procedure to generate a hypothesis from an initial state assignment, we will refer to such procedure as Hypothesis. The operation Hypothesis accepts as an input a hypothetical state assignment PH over a causally independent set $V' \subseteq V$ and returns a hypothetical state H_t. The procedure Hypothesis is defined in the following algorithm.

The Hypothesis procedure described in Algorithm 1 is not deterministic, as it will generate different H depending on the selection from PSet. However, it is easy to see that this algorithm will terminate after a finite number of steps. Clearly, the generated H is complete, namely defined for all $v \in V$. Moreover we have the following result.

Proposition 4 Let $V' \subseteq V$ be a causally independent set and PH be consistent hypothetical assignment for V. Then any hypothetical state H output from Hypothesis(V', PH) is consistent and maximal with respect to PH.

Proof (Outline): We already shown that recursive application of the BProp and FProp operators generate a maximal consistent set with respect to the initial state assignment PH. Let $V'' = V_c$ and $PH'' = H$ just after the termination of the recursive BProp and FProp operations. It is easy to see that V'' does not contain any descendants from $V - V''$. In addition, given that PH'' is maximally consistent, every node in V'' is either explained by another node in V'' or is a root node (failure mode) in the sTFPG mode. Therefore, assigning an OFF state to all the nodes in $V - V''$ will preserve the consistency and maximality of the state assignment. □

The above results simply says that a consistent state assignment over a causally independent set is a valid state assignment, as it can always be extended to a (complete) hypothesis. It is important to note that the number of choices available in

Algorithm 1 The Hypothesis procedure

input: $V' \subseteq V$ and $PH_t : V' \rightarrow \{\texttt{ON}, \texttt{OFF}\} \times \mathbb{R} \times \mathbb{R}$
$V_c = V'$
$H = PH_t$
define $\mathsf{In}(X) := \{v \in X | (\forall v' \in X)\ (v, v') \notin E\}$
define $\mathsf{PSet}(X) := \{v \in V - X | (\exists v' \in \mathsf{In}(X))\ (v, v') \in E\}$
while $\mathsf{PSet}(V_c) \neq \varnothing$ **do**
  select v from $\mathsf{PSet}(V_c)$
  $H = \mathsf{BProp}(H, V_c, v)$
  $V_c = V_c \cup \{v\}$
end while
define $\mathsf{ODC}(X) := \cup_{v \in X} \mathsf{Reach}(v) - X$
define $\mathsf{TSet}(X) := \{v \in V - X | \mathsf{ODC}(X) \times v \cap E = \varnothing\}$
define $\mathsf{CSet}(X) := \{v \in \mathsf{TSet}(X) | (\exists v' \in X)\ (v', v) \in E\}$
while $\mathsf{CSet}(V_c) \neq \varnothing$ **do**
  select v from $\mathsf{CSet}(V_c)$
  $H = \mathsf{FProp}(H, V_c, v)$
  $V_c = V_c \cup \{v\}$
end while
for all $v \in V - V_c$ **do**
  $H(v).\text{state} = \mathsf{OFF}$
  $H(v).\text{terl} = H(v).\text{terl} = 0$
end for
return H

the above Hypothesis procedure is finite and therefore, the number of all different hypotheses that can be generated for a given V', PH is finite.

4.2 Hypothetical state initialization

The previous section shows that it is possible to generate a hypothesis from a consistent state assignment over a set of causally independent nodes. To generate a diagnosis, the underlying hypothesis must to be of the highest possible rank, namely, maximally matching with the current observed state. Clearly the rank of the hypothesis is dependent on the initial state assignment.

Intuitively, a valid initial set that matches the current observed state is favorable, as such a set can lead to a high ranking hypothesis. The support set Ψ_t discussed earlier provides a collection of maximal consistent discrepancy states. The problem, however, is that the underlying sets may not be causally independent and, therefore, there is no guarantee that they can be extended to a consistent hypothetical state.

To generate a valid initial state assignment we need to add the condition of causal independence to the timing consistency requirement for the state assignment. To this end, a set $D' \subseteq D$ is said to be strongly T-consistent, or ST-consistent for short, if D' is both T-consistent and causally independent.

We define the *consistency index* at time t, denoted, μ_t to be the size of the largest possible ST-consistent set. The *evidence set* at time t, is the set of all ST-consistent sets with size μ_t. This set will be denoted Ω_t. Similar to the the set Ψ_t, the set Ω_t can be computed incrementally by maintaining a list of all ST-consistent sets, and updating this set when a new sensor signal is received.

Based on the above definitions we can have the following result for the maximum hypothesis rank, R_t, at time t. Recall, that λ_t refers to the size of the maximal T-consistent set.

Proposition 5

$$\mu_t \leq R_t \leq \lambda_t$$

Proof (Outline): For the lower bound, consider an initial state assignment that strongly matches a maximal set of ST-consistent discrepancies $\omega \in \Omega_t$. By definition this assignment is a valid one and, therefore, can be extended to a complete hypothesis, H, which preserves the state of the initial assignment. The rank of H must be greater or equal to μ_t in this case. As H will be strongly matching with the set ω.
For the upper bound assume there exists a hypothesis with a rank $r > \lambda_t$. This means that it can match the state of r discrepancies. It can be shown from the definition of hypothesis consistency and T-consistent set that such set of discrepancies must be T-consistent. This contradicts that λ_t is the largest size of T-consistent sets at the current time t. □

The above result have several important consequences. The first is that, the best possible initial state assignment should be a one that exactly matches the observed states of any discrepancy set in Ω_t. Another important implication is that if $\mu_t = \lambda_t$ then the optimal diagnosis can be obtained directly applying the Hypothesis operation to each set in Ω_t. Note that the value of λ_t and μ_t are time dependent and they reflect the existence of faulty sensors (as they introduce inconsistency between observed states).

4.3 Generating the diagnosis set

Initially, at the start, $t = 0$, we have $R_o = N$ (N is the number of discrepancies in the model.) Given that consistency between nodes can only be invalidated with new sensor signals, the maximum rank R_t is a monotonically decreasing function of time. Clearly, if there is no sensor failure (more precisely, inconsistency between observed discrepancy states), R_t will remain at its maximum value N. The following result shows that it is possible to compute the optimal hypothesis for any sTFPG model with a given observed state.

Proposition 6 For a given sTFPG and observed state S_t the problem of finding the diagnosis set at time t is decidable.

Proof (Outline): We show early that the set Ω_t is finite and the number of hypotheses generated for each state assignment is also finite. Therefore, the optimal set of hypotheses can be obtained by running the procedure Hypothesis finite number of times for each element in Ω_t. □

Because of the complex nature of the consistency conditions, there does not seem to be a direct way to generate an optimal diagnosis starting from an initial assignment from the set Ω_t. However, it is possible to reduce the search for optimal hypothesis by eliminating those paths that could not improve the current maximal ranking.

At the occurrence of every event (sensor signal or timeout) a failure report is generated from the set of optimal hypotheses. This report consists of the set of all consistent state assignments that maximally matches the current set of observation. In this report, any observed state that does not (weakly) match the current hypothesis is considered faulty.

Faulty alarms with observed ON state are considered false alarms, while those with OFF state are considered missing alarms.

5 Conclusion

In this paper we introduced an consistency-based approach for robust diagnosis of temporal causal systems based on the simple timed failure propagation graph model. The model can be used for diagnosis a general class of systems with temporal propagation constraints. We presented the main elements of the diagnostic problem and formally described the main parts of the optimal diagnosis reasoning algorithm.

In future work, we plan to extend the result to the more general hybrid failure propagation graph model that allows dependency loops, AND-type discrepancies, and mode switching.

References

[Abdelwahed *et al.*, 2004] S. Abdelwahed, G. Karsai, and G. Biswas. System diagnosis using hybrid failure propagation graphs. In *The 15th International Workshop on Principles of Diagnosis*, Carcassonne, France, 2004.

[Console and Torasso, 1991] L. Console and P. Torasso. On the co-operation between abductive and temporal reasoning in medical diagnosis. *Artificial Intelligence in Medicine*, 3(6):291–311, 1991.

[Darwiche and Provan, 1996] A. Darwiche and G. Provan. Exploiting system structure in model-based diagnosis of discrete-event systems. In *Proc. of the Seventh International Workshop on Principles of Diagnosis*, pages 95–105, 1996.

[Darwiche, 1998] A. Darwiche. Model-based diagnosis using structured system descriptions. *Journal of Artificial Intelligence Research*, 8:165–222, 1998.

[de Kleer *et al.*, 1992] J. de Kleer, A. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56, 1992.

[Frank, 1990] P. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica*, 26(3):459–474, 1990.

[Gamper, 1996] J. Gamper. *A Temporal Reasoning and Abstraction Framework for Model-Based Diagnosis Systems*. PhD thesis, RWTH, Aachen, Germany, 1996.

[Hamscher *et al.*, 1992] W. Hamscher, L. Console, and J. de Kleer. *Readings in Model-Based Diagnosis*. Morgan Kaufmann Publishers, 1992.

[Hessian *et al.*, 1990] R. Hessian, B. Salter, and E. Goodwin. Fault-tree analysis for system design, development, modification, and verification. *IEEE Transactions on Reliability*, 39(1):87–91, 1990.

[Himmelblau, 1978] D. M. Himmelblau. Fault detection and diagnosis in chemical and petrochemical processes. *Chemical Eng. Mon.*, 8, 1978.

[Ishida *et al.*, 1985] Y. Ishida, N. Adachi, and H. Tokumaru. Topological approach to failure diagnosis of large-scale systems. *IEEE Trans. Syst., Man and Cybernetics*, 15(3):327–333, 1985.

[Karsai *et al.*, 1992] G. Karsai, J. Sztipanovits, S. Padalkar, and C. Biegl. Model based intelligent process control for cogenerator plants. *Journal of Parallel and Distributed Systems*, 15:90–103, 1992.

[Karsai *et al.*, 2003] G. Karsai, G. Biswas, and S. Abdelwahed. Towards fault-adaptive control of complex dynamic systems. In T. Samad and G. Balas, editors, *Software-Enabled Control: Information Technology for Dynamical Systems*, chapter 17. IEEE publication, 2003.

[Misra *et al.*, 1994] A. Misra, J. Sztipanovits, and J. Carnes. Robust diagnostics: Structural redundancy approach. In *SPIE's Symposium on Intelligent Systems*, 1994.

[Misra, 1994] A. Misra. *Sensor-Based Diagnosis of Dynamical Systems*. PhD thesis, Vanderbilt University, 1994.

[Mosterman and Biswas, 1999] P. J. Mosterman and G. Biswas. Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans. on Systems, Man and Cybernetics*, 29(6):554–565, 1999.

[Padalkar *et al.*, 1991] S. Padalkar, J. Sztipanovits, G. Karsai, N. Miyasaka, and K. C. Okuda. Real-time fault diagnostics. *IEEE Expert*, 6(3):75–85, 1991.

[Patton, 1994] R. Patton. Robust model-based fault diagnosis:the state of the art. In *IFAC Fault Detection, Supervision and Safety for Technical Processes*, pages 1–24, Espoo, Finland, 1994.

[Rao and Viswanadham, 1987a] S. V. Nageswara Rao and N. Viswanadham. Fault diagnosis in dynamical systems: A graph theoretic approach. *Int. J. Systems Sci.*, 18(4):687–695, 1987.

[Rao and Viswanadham, 1987b] S. V. Nageswara Rao and N. Viswanadham. A methodology for knowledge acquisition and reasoning in failure analysis of systems. *IEEE Trans. Syst., Man and Cybernetics*, 17(2):274–288, 1987.

[Reiter, 1987] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.

[Richman and Bowden, 1985] J. Richman and K. R. Bowden. The modern fault dictionary. In *International Test Conference*, pages 696–702, 1985.

[Scherer and White, 1989] W. T. Scherer and C. C. White. A survey of expert systems for equipment maintenace and diagnostics. In S. G. Tzafestas, editor, *Knowledge-based system diagnosis, supervision and control*, pages 285–300. Plenum, New York, 1989.

[Tzafestas and Watanabe, 1990] S. Tzafestas and K. Watanabe. Modern approaches to system/sensor fault detection and diagnosis. *J. A. IRCU Lab.*, 31(4):42–57, 1990.

[Viswanadham and Johnson, 1988] N. Viswanadham and T. L. Johnson. Fault detection and diagnosis of automated manufacturing systems. In *27th IEEE Conference on Decision and Control*, 1988.