

Notions of Diagnosability for Timed Failure Propagation Graphs

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

Abstract—Timed failure propagation graphs (TFPG) are causal models that capture the temporal aspects of failure propagation in dynamic systems. This paper presents several notions of diagnosability for timed failure propagation models. Diagnosability characteristics of TFPG models are defined based on three metrics; failure detectability, distinguishability, and predictability. The paper identifies the modeling and run-time parameters of the diagnosability metrics. The application of diagnosability analysis to the problem of alarm allocation is discussed.

I. INTRODUCTION

Timed failure propagation graphs (TFPG) [1], [2] are causal models that describe the system behavior in presence of faults. The TFPG structure captures the effect of timing constraints and switching dynamics on the propagation of failures in practical discrete event and hybrid systems. The TFPG model is closely related to the fault model presented in [3], [4] and used for an integrated fault diagnoses and process control system. The temporal aspects of the TFPG model is closely related to the domain theoretic notion of temporal dependency proposed in [5]. However, TFPG based diagnosis is an online incremental approach that focuses on diagnosis robustness with respect to alarm failure. In addition, we consider a specific form of temporal dependency that directly incorporate the effect multi-mode dynamics on failure propagation.

In [6] we presented a consistency-based approach for robust diagnosis of systems in which failure behavior can be captured by TFPG models. This proposed diagnosis approach is conducted online based on incremental non-monotonic reasoning that is robust with respect to various forms of alarm failures. The proposed algorithm consists of two main procedures; the first one generates an optimal consistent initial state assignment based on current state observation while the second procedure generates the set of all consistent hypothesis from a given initial state assignment.

In this paper, we introduce several notions of diagnosability for the timed failure propagation model. Diagnosability of a system, in its most general sense, means that property of the system which allows the faults to be detected and diagnosed in a timely manner, that is, within a finite interval from the time at which the failure occur. In this paper, three diagnosability metrics are defined, namely, failure detectability, distinguishability, and predictability. The paper presents the formal definitions as well as the basic algorithms for evaluating these metrics for TFPG models.

Diagnosability analysis have been addressed in literature for various diagnosis models. Diagnosability and optimal alarm allocation for linear discrete dynamic systems has been investigated in [7]. In this work indices for detectability of a fault and separability of faults is incorporated in the state-space representation of a system. In [8] diagnostic dictionaries are used to model the failures and computes the t -fault diagnosability (up to t faults can be diagnosed correctly or not) of the faults in a system. A diagnosability criterion, which expresses the permissible ambiguity in the diagnosis, is defined for every component in the system. In [9] an approach to evaluating alarm placement is presented. They simulate the system behavior using analytical models to predict the signal values read by alarms under normal and faulty operation.

Diagnosability of discrete event systems models has been originally addressed in [10]. In this work, diagnosability is defined from a formal language perspectives as the ability to detect failure state without ambiguity after the occurrence of a finite number of observable event. This diagnosability notation has been extended to timed automata in [11].

The paper is organized as follows. In Section 2, the timed failure propagation graph model is introduced. Section 3 presents an overview of the consistency based diagnosis approach for TFPG. Section 4 introduces three diagnosability metrics; failure detectability, distinguishability, and predictability, in the context of TFPG models. Section 5 discusses methods for evaluating TFPG diagnosability based on the proposed metrics. In Section 6, we discuss the application of diagnosability analysis to the alarm allocation problem. Conclusion and future works are presented in Section 7.

II. TIMED FAILURE PROPAGATION GRAPHS

A TFPG 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. To represent of failure propagation in multi-mode (switching) systems, edges in the graph model can be activated or deactivated depending on a set of possible operation modes of the system. Formally, a TFPG is represented as a tuple $G = (F, D, E, M, ET, EM, DC)$, where:

- F is a nonempty set of failure nodes.
- D is a nonempty set of discrepancy nodes.

- $E \subseteq V \times V$ is a set of edges connecting the set of all nodes $V = F \cup D$. $src(e)$ and $dst(e)$ denotes the source and destination nodes of the edge e , respectively.
- M is a nonempty set of system modes. At each time instance t the system can be in only one mode.
- $ET : E \rightarrow Int$ is a map that associates every edge in E with a time interval.
- $EM : E \rightarrow \mathcal{P}(M)$ is a map that associates every edge in E with a set of modes in M . We assume that $EM(e) \neq \emptyset$ for any edge $e \in E$.
- $DC : D \rightarrow \{AND, OR\}$ is a map defining the class of each discrepancy as either AND or an OR node.

The set V contains $n + m$ vertices, representing n failure modes and m discrepancies. The map ET associates each edge $e \in E$ with the minimum and maximum time for the failure to propagate along the edge. For an edge $e \in E$, we will use the notation $e.tmin$ and $e.tmax$ to indicate the corresponding minimum and maximum time for failure propagation along e , respectively. That is, given that a propagation edge is enabled (active), it will take at least (most) $tmin$ ($tmax$) time for the fault to propagate from the source node to the destination node. The map EM associates each edge $e \in E$ with a subset of the system modes at which the failure can propagate along the edge. Consequently, the propagation link e is enabled (active) in a mode $m \in M$ if and only if $m \in EM(e)$. The map DC defines the type of a given discrepancy as either AND or OR. An OR type discrepancy node will be activated when the failure propagate to the node from any of its parents. On the other hand, an AND discrepancy node can only be activated if the failure propagates to the node from all its parents. We assume that TFGP models do not contain self loops and that failure modes are always root nodes, i.e., they cannot be a destination of any edge. Also, a discrepancy cannot be a root node, that is, every discrepancy must be a successor of another discrepancy or a failure mode.

Figure 4 shows an example of failure propagation graph model. Rectangles in the graph model represent the failure modes while circles and squares represent OR and AND type discrepancies, respectively. The arrows between the nodes represent failure propagation. Propagation edges are parameterized with the corresponding interval, $[e.tmin, e.tmax]$, and the set of modes at which the edge is active. The above figure shows also a sequence of alarm signals identified by shaded discrepancies. The time at which the alarm is observed is shown above the corresponding discrepancy.

The TFGP model captures observable failure propagations between discrepancies in practical systems. In this setting, alarms capture state deviations from nominal values. The set of all observed deviations corresponds to the discrepancy set in the TFGP model. Propagation edges, on the other hand, correspond to causality (as defined by energy flow, for instance) 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 depends on the system's time constants as well as the size and timing of underlying failure. Propagation delay intervals can be

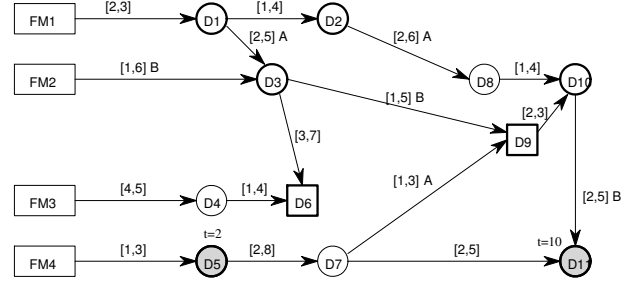


Fig. 1. A timed failure propagation graph

computed analytically or through simulation of an accurate model.

Failure propagation in a TFGP 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, if it is an OR type node. Failure can only reach nodes of AND type only if it already reached all its parents. For an OR type node v' and an edge $e = (v, v') \in E$, once a failure effect reaches v at time t it must reach v' at a time t' where $e.tmin \leq t' - t \leq e.tmax$. On the other hand, the activation period of an AND alarm v' is the min/max composition of the activation periods for each link $(v, v') \in E$. For a failure to propagate through a link (v, v') , the link should be active throughout the propagation, that is, from the time the failure reaches v to the time it reaches v' . If the link is deactivated any time during the propagation (because of mode switching), the propagation stops. Links are assumed memoryless with respect to failure propagation so that current failure propagation is independent of any (incomplete) previous propagation. Also, once a failure effect reaches a node its state will change permanently, and it will not be affected by any future failure propagation.

III. REVIEW OF THE DIAGNOSIS APPROACH

An *actual (physical) system* state corresponds to the current state of all nodes in the TFGP model. A physical state a time t is given by a map $AS_t : V \rightarrow \{ON, OFF\} \times \mathbb{R}$, where V is the set of nodes in the TFGP model. An ON state for a node indicates that the failure (effect) reached this node, otherwise it is set to OFF. The state at time t is denoted $AS_t(v).state$, while $AS_t(v).time$ denote the last time at which the state of v is changed. Failure effects are assumed permanent, therefore, the state of a node once changed will remain constant after that. Due to mode switching, a similar map is used to define the state of edges, namely, $ES_t : E \rightarrow \{ON, OFF\} \times \mathbb{R}$, with $ES_t(e).state$ defines the state of the edge e at time t , while $ES_t(e).time$ is the last time at which the state of e is changed.

The observed state of the system may not be consistent with the failure propagation graph model temporal constraints, due to potential alarm failures. The *observed state* at time

t is defined as a map $S_t : D \rightarrow \{\text{ON}, \text{OFF}\} \times \mathbb{R}$. Clearly, observed states are only defined for discrepancies. We assume that alarm signals are permanent so that the observed state of a discrepancy once changed will remain constant after that. This assumption of permanent change also applies to faulty alarms.

The aim of the diagnosis reasoning process is to find a consistent and plausible explanation of the current system state based on the observed state. Such explanation is given in form of a valid hypothetical state. A *hypothetical state* is a map that defines node states and the interval at which each node changes its state. Formally a hypothetical state at time t is a map $H_t : V \rightarrow \{\text{ON}, \text{OFF}\} \times \mathbb{R} \times \mathbb{R}$. Similar to actual states, hypothetical states are defined for both discrepancies and failure modes. The estimated earliest (latest) time of state change is denoted $H(v).\text{terl}$ ($H(v).\text{tlat}$).

A hypothetical state is 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 is valid only if it is consistent with the TFPG model. State consistency in TFPG models is a node-parents relationship that can be extended pairwise to arbitrary subsets of nodes. State consistency can be defined based on the semantics of failure propagation, can be used to check if an OR (AND) alarm is consistent with any of (all) its parent. Because of mode switching, the node-parents consistency depends not only on the state of the underlying nodes but also on the current state of the connection edges and their last activation times. Formally, let $d \in D$ be an OR type discrepancy. Then, a hypothesis map H_t is *consistent* with respect to d if:

- 1) $H_t(d) = \text{OFF}$ and for all $(v, d) \in E$:
 - a) $H_t(v) = \text{OFF}$, or
 - b) $H_t(v) = \text{ON} \wedge ES_t(v, d).\text{state} = \text{ON} \wedge t < \max(H_t(v).\text{tlat}, ES_t(e).\text{time}) + (v, d).\text{tmax}$
- 2) $H_t(d) = \text{ON}$ 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) $H_t(d).\text{tlat} \leq \min_{v \in U_d} \{H_t(v).\text{tlat} + (v, d).\text{tmax}\}$
$$U_d = \{v \in V \mid (v, d) \in E \text{ and } H_t(v).\text{state} = \text{ON}\}$$

The consistency relationship is the foundation of the diagnosis approach presented in [6]. This relationship is used to check the consistency of observed discrepancies. It is also used to generate a maximal (with respect to activation period) hypothetical state for an alarm given the hypothetical state of its parents. Ultimately, this can be used to extend a hypothetical state defined for a subset $V' \subseteq V$ through forward-propagation up to the leaf nodes. Extension through backward-propagation of hypothetical states can also be defined based on the consistency relationship. The optimal diagnosis algorithm¹ consists of the following main steps:

- Compute the maximal set(s) of consistent discrepancy with respect to the current observed state.
- Generate a consistent partial hypothetical states covering the observed discrepancy.

- Propagate the generated partial hypothetical states backward up to the parent set of failure mode. All other failure modes are assigned the default OFF state.
- Propagate the new partial hypothetical state forward recursively until all nodes in the TFPG model are covered.

This procedure is conducted incrementally based on previous computation at the occurrence of every state-changing event (alarm signal or time-out). A failure report is then generated from the computed set of optimal hypotheses. The failure report enlists the set of all consistent state assignments that maximally matches the current set of observation. Any observed state that does not match the current hypothesis is considered faulty. A detailed description and analysis of the diagnosis algorithm can be found in [6].

IV. DIAGNOSABILITY METRICS

Diagnosability is a property of the system that allows faults to be detected and identified in a timely manner. Diagnosability of a system depends on how the system is modeled as well as on the outcome of the diagnosis algorithm (particular, on being optimal or sub-optimal). In order to characterize the diagnosability of a system, one needs to develop some criteria or metrics, which express the property of diagnosability in a reasonable and coherent manner. In this section, we outline three traditional metrics for diagnosability, namely, fault detectability, distinguishability, and predictability, and define them within the context of the timed failure propagation graph models. The rationale for these definitions comes from the fundamental concepts of diagnosis and are independent of the actual reasoning method.

A. Fault Detectability

In general, the presence of a fault in a system is detected by observation of one or more discrepancies. Because of the complex relationship between faults and discrepancies, more than one fault may be suspected when a discrepancy is observed. This set represents the ambiguity in the diagnostic results at any given time. Because of the dynamics of the system, some time may pass before a fault causes the observed discrepancies. The fault may propagate to more discrepancies, but the evidence provided by these additional discrepancies will only contribute to isolating the fault.

The fault is said to be detected at the time it is first included in the ambiguity set. Note that, a failure may be removed from the ambiguity set after being included. This is mainly due to the fact that failure detection (inclusion in the ambiguity set) depends on certain assumptions about the operating conditions. We will discuss such conditions in more details later. For now we just use C to denote a given set of operating conditions.

Detectability of a failure mode $f \in F$, with respect to operating conditions C , is denoted by $\text{DET}_C(f)$ and is given as a tuple $(\text{tmin}, \text{tms}, \text{tmax})$, where tmin is the minimum time that will pass before f can be included in the ambiguity set after its occurrence, tms is the minimum time that will pass before f is stabilized in the ambiguity set (that is, ambiguity sets generated after $\text{DET}_C(f).\text{tms}$ will always contain f), and

¹Optimality here is defined with respect to the size of the ambiguity set

t_{\max} is the maximum time that will pass before its inclusion in the ambiguity set. We will drop C as a subscript if it is clear from the context.

Detectability of a single failure can be extended to a set of failures $F' \subseteq F$. The definition of detectability for a group of failures is a simple extension of the single fault case. For simplicity, we will use the same notation for a set of failure modes and write $\text{DET}(F')$ for the tuple $\langle t_{\min}, t_{\text{ms}}, t_{\max} \rangle$, where t_{\min} is the minimum time that will pass before the entire set F' can be included in the ambiguity set after their simultaneous occurrence, and so on. The computation of $\text{DET}(F')$ is not a direct composition of the underlying $\text{DET}(f)$, $f \in F'$ due to the potential effect of AND discrepancies.

B. Fault Distinguishability

Once a fault has been detected, it needs to be isolated from other possible faults. A diagnoser does this by pruning the elements of ambiguity set using consistency-based reasoning techniques down to the actual fault. During the diagnosis process, more evidence may come in causing the ambiguity set to shrink or grow. The goal is that at the end of diagnosis, the ambiguity set should contain only the fault that actually occurred, i.e., there should be no ambiguity in the diagnosis. In case of multiple faults, the set should contain exclusively all the faults that actually occurred.

The time needed for the resolution of ambiguity between faults may vary depending on the dynamics of the system and alarm coverage which determines the available evidence at a given time. Similar to detectability, failure distinguishability depends also on several run time conditions. Therefore, the distinguishability metrics need to be addressed with respect to a set of assumptions C defining the run-time operating conditions.

Distinguishability of a failure mode $f \in F$ is denoted $\text{DST}(f)$ and is given as a tuple $\langle t_{\min}, t_{\text{ms}}, t_{\max} \rangle$, where $\text{DST}(f).t_{\min}$ ($\text{DST}(f).t_{\max}$) is the minimum (maximum) time that will pass before f is the only element in the ambiguity set after its occurrence, and $\text{DST}(f).t_{\text{ms}}$ is the minimum time at which f is stabilized as the only element in the ambiguity set. Similar to detectability, the distinguishability of a single failure can be extended to a set of failures $F' \subseteq F$.

C. Fault Predictability

In many practical situations, not only the fault but also its consequences are of major concerns. Fault propagation is the mechanism which may trigger a discrepancy in the system behavior with critical consequences. A task related to diagnosis is the timely prediction of critical failures that can take place. This task becomes essential in order to predict potentially catastrophic failures and take preventive measures.

Predictability of a discrepancy $d \in D$, denoted by $\text{PRD}(d)$, is the interval (t_{\min}, t_{\max}) , where $\text{PRD}(d).t_{\min}$ is the shortest available time period between a forewarning and the actual occurrence of d , and $\text{PRD}(d).t_{\max}$ is the longest available time period between a forewarning and the actual

occurrence of d . If $\text{PRD}(d).t_{\min} = \text{PRD}(d).t_{\max} = 0$, then d is not predictable. Clearly, also here both $\text{PRD}(d).t_{\min}$ and $\text{PRD}(d).t_{\max}$ depend on the operating conditions of the systems and will therefore be evaluated with respect to a given set of assumptions C about these conditions.

V. DIAGNOSABILITY ANALYSIS

As mentioned earlier, diagnosability metrics depend on the system model as well as certain run-time conditions. In particular, two factors related to the TFPG model have a direct effect on diagnosability, namely, the number of alarms and their location, i.e., the physical layout of alarm distribution in the TFPG model. There are three main run-time conditions that directly influence the diagnosability metrics, namely, alarm failures, the timing characteristics of mode switching as well as that of failure modes. In particular, diagnosability of TFPG models are evaluated, in general, with respect to the following:

- The maximum number of alarm failures, $N \in \mathbb{N}_+$. Alarm failures include both missing alarms (inconsistently OFF) and false alarms (inconsistently ON).
- The timing characteristics of the mode switching. Such characteristics provide the information needed to identify all valid mode-switching (event) sequences. A possible characterization can be given as a tuple $T = (t_{\min}, t_{\max}, tp)$, where $T.t_{\min}$ and $T.t_{\max}$ are the minimum and maximum time between mode switching, respectively and $T.tp$ is the average period between mode-switching events. Accordingly, if the system switch to a new mode at time t , the next mode switching should occur before $t + T.t_{\min}$ and not later than $t + T.t_{\max}$. Mode-switching can also be captured as a stochastic event in which case the period between two mode-switching events is estimated using a probability distribution function. Using this form of representation, however, will require redefining diagnosability metrics as probability measures.
- In the case of multiple failure modes, the relative time between failure events directly influence the first two diagnosability metrics. This time can be characterized using either boundary or probability measures similar to the case of mode-switching events.

The effect of mode-switching on system diagnosability is complicated to handle both analytically and computationally. In this paper, we will consider a simple representation of mode switching effect in which additional additional delay period is added to each propagation link representing average situations under mode-switching. Hereafter, we assume that the propagation delays in the model are adjusted to take into account the effect of mode-switching and therefore we will only consider the effect of alarm failure. Also, we will only consider diagnosability measures with respect to single failure. The extension to a set of failure modes is laborious but straightforward.

Considering first the case of failure detectability. Let $f \in F$ be a failure mode in TFPG model and N be the maximum

number of alarm failures that may occur after f . A set of run-time operation scenarios can be defined given N as maximum possible number of alarm failures. An operation scenario is a sequence $R_a = (e_{d1}, e_{d2}, \dots, e_{dM})$ of the alarm failure events $e_{di} = (d_i, s_i, t_i)$ with $d_i \in D$ is a failed alarm, s_i is the state at which the alarm fail (either being permanently ON or permanently OFF), and $t_i > 0$ is the time at which d_i failed with $t_i \geq t_{i+1}$ and $M \leq N$. We will write \mathcal{R}_C to denote the set of all valid scenarios².

Assuming that a failure mode f is triggered at time $t = 0$. The interval during which the attached discrepancies is triggered depends only on a given scenario $R \in \mathcal{R}_C$. Based on the semantics of failure propagation in TFPG, one can define precisely the activation intervals for the set D (given f occurring at time $t = 0$ and a scenario R) as a map $\text{Act}_f^R : D \rightarrow \text{Int}$ where Int is the set of all intervals over \mathbb{R} . The minimum detection time, $\text{DET}(f).\text{tmin}$, can then be defined as

$$\min_{R \in \mathcal{R}_C} \{\text{Act}_f^R(d).\text{tmin} \mid (f, d) \in E, \text{DC}(f) = \text{OR}\}$$

The above formula can be used directly to compute $\text{DET}(f).\text{tmin}$ as the best case condition does not involve any failed alarms. The maximum time to detect f , $\text{DET}(f).\text{tmax}$, on the other hand, corresponds to the worst case scenario where the detection of f is delayed (as a result for failed alarms) to the maximum limit. Consider for instance, the case where $N = 1$. Clearly, in this case, the detection of f will be delayed if the first alarm to trigger in response to f is faulty (will not switch to ON at the right time). In this case, f can only be detected when the second alarm in the sequence is triggered. Accordingly, $\text{DET}(f).\text{tmax}$ corresponds to the maximum time for the second alarm to trigger, considering the delay encountered from mode-switching. In the general case, we have the following result.

Proposition 1: Let $D_t^f \subseteq D$ be the set of alarms that are triggered at or before t after the occurrence of f . Assume that N alarms are faulty. Then f will always be part of the ambiguity set at any time t' if $|D_t^f| \geq 2N$.

Proof (Outline): This can be proved by induction. Considering the base case $N = 1$. Clearly, the worst case situation corresponds to the case when the first alarm is a (missing) faulty one, in which case f will not be detected. The next alarm is a valid alarm and will then allow f to be detected (there will be two plausible hypotheses in this case, and one of them will point to f). Now assuming the proposition is valid for N . Then f can be detected at the time the $2N$ th alarm is triggered. Adding two alarms to this case with one of them faulty will also allow f to be detected as the evidence for f remains balanced w.r.t to the evidence against it. This is based on the fact that the optimal algorithm [6] always generate the

²To add mode switching with timing characteristics $T = (\text{tmin}, \text{tmax}, \text{tp})$, we will need to consider another independent set of mode switching event sequences, each given as $R_m = (e_{m1}, e_{m2}, \dots, e_{mK})$ where $e_{mi} = (m_i, t_i)$ with $m_i \in M$ is the mode the system switches to at time t_i and satisfying $T.\text{tmin} \leq t_{i+1} - t_i \leq T.\text{tmax}$. In this case a scenario is a tuple (R_a, R_m) .

maximum consistent hypothetical state containing the current observation. \square

Based on the above result, $\text{DET}(f).\text{tmax}$ corresponds to the time at which the $2N$ th alarm is triggered assuming a scenario R without any alarm failures and taking into account the delay caused by mode-switching. The following algorithm computes $\text{DET}(f).\text{tmax}$ for a given number of maximum alarm failure N .

Algorithm 1 Computing $\text{DET}(f).\text{tmax}$

```

input:  $f \in F, N$ 
 $D := \emptyset; \text{FTMAX} := -1$ 
 $\text{DSet} := \{d \in D \mid (f, d) \in E, \text{DS}(d) = \text{OR}\}$ 
 $T : D \cup \{f\} \rightarrow \mathbb{R} : v \mapsto -1$ 
 $H : D \cup \{f\} \rightarrow \{\text{T}, \text{F}\} : d \mapsto \text{F}$ 
 $T(f) := 0; H(f) := \text{T}$ 
while  $\text{DSet} \neq \emptyset$  do
   $d = \text{pop}(\text{DSet})$ 
   $H(d) = \text{True}$ 
  for  $(v, d) \in E$  do
    if  $v \in \text{Reach}(f)$  and  $H(v) = \text{F}$  then  $H(d) = \text{F}$ 
    if  $T(v) \neq -1$  then
       $T(d) = \max\{T(d), T(v) + (v, d).\text{tmax}\}$ 
      if  $T(d) > \text{FTMAX}$  then  $\text{FTMAX} = T(d)$ 
    end if
  end for
  for  $(d, v) \in E$  do
    if  $v \notin \text{DSet}$  then  $\text{push}(v, \text{DSet})$ 
  end for
  if  $H(d) = \text{F}$  then  $\text{push}(d, \text{DSet})$ 
end while
return  $\text{FTMAX}$ 

```

Note that, for a single failure mode, AND alarms are irrelevant to the computation of $\text{DET}(f).\text{tmax}$ for a single failure mode. Computing $\text{DET}(f).\text{tms}$ is similar to that of $\text{DET}(f).\text{tmax}$ given that f can only stabilize in the ambiguity set after the triggering $2N$ alarms in response to f (with no alarm failures). In this case, it is only required to use the minimum propagation tmin time instead of tmax in the above algorithm.

Computing the distinguishability metrics is conceptually similar noting that optimal algorithm will produce f as the only element of the ambiguity if and only if the number of evidence supporting f is greater than the number of evidence against it. In the computation of the metrics, this translates into the number of observed healthy alarms vs. the number of failed ones.

VI. APPLICATION TO ALARM ALLOCATION

As discussed earlier, diagnosability of the system depends directly on the alarm distribution topology as well as the operating conditions. Given that operating conditions are typically uncontrollable, one can only improve the system diagnosability by changing the alarm distribution topology, namely, the number of alarms and their layout. Although it

is theoretically possible to add more alarms to improve the system diagnosability, it is not always possible practically to put an alarm on every important physical variable in the system. Given a limited number of alarm, it is important to evaluate the relative importance of alarms from the point of view of diagnosability. Such evaluation can be used to choose the best allocation of alarms to insure proper diagnosability of the system while keeping costs at a minimum.

The Alarm allocation problem typically starts with a partial alarm allocation in which a subset of alarms that are already part of the system design and other discrepancies are left open for possible alarm attachment. Diagnosability analysis is then used to allocate additional alarms and remove or redistribute some of them. Such analysis generates suggestions for alarm allocation by specifying the diagnosability criteria that is required to be met. The diagnosability criteria consists of sets of failure modes and discrepancies and their desired detectability, distinguishability and predictability metrics. Given that there is finite number of possible alarm allocations, such procedure is computationally feasible albeit expensive.

VII. CONCLUSION

In this paper we introduced a set of diagnosability notions for temporal causal systems modeled using timed failure propagation graph model. Due to the complexity added by incorporating timing information, mode-switching, and potential alarm failure to the TFPG semantics, diagnosability metrics depends not only on the topology of the failure model but also on how these new aspects can be effectively characterized for typical run-time scenarios. This paper presents an initial attempt to characterize the factors affection diagnosability in a computationally feasible way. The paper further discusses the application of diagnosability analysis to the alarm allocation problem.

In future work, we plan to extend the proposed diagnosability metrics to a more general framework that include direct measures for mode-switching and failure mode sequences. We are currently working to implement the diagnosability analysis methods described in this paper in a failure model analysis tool that can provide feedback on alarm allocation based on a set of diagnosability preferences.

REFERENCES

- [1] A. Misra, "Sensor-based diagnosis of dynamical systems," Ph.D. dissertation, Vanderbilt University, 1994.
- [2] A. Misra, J. Sztipanovits, and J. Carnes, "Robust diagnostics: Structural redundancy approach," in *SPIE's Symposium on Intelligent Systems*, 1994.
- [3] S. Padalkar, J. Sztipanovits, G. Karsai, N. Miyasaka, and K. C. Okuda, "Real-time fault diagnostics," *IEEE Expert*, vol. 6, no. 3, pp. 75–85, 1991.
- [4] G. Karsai, J. Sztipanovits, S. Padalkar, and C. Biegl, "Model based intelligent process control for cogenerator plants," *Journal of Parallel and Distributed Systems*, vol. 15, pp. 90–103, 1992.
- [5] V. Brusoni, L. Console, P. Terenziani, and D. T. Dupre, "A spectrum of definitions for temporal model-based diagnosis," *Artificial Intelligence*, vol. 102, no. 1, pp. 39–79, 1998.
- [6] S. Abdelwahed, G. Karsai, and G. Biswas, "A consistency-based robust diagnosis approach for temporal causal systems," in *The 16th International Workshop on Principles of Diagnosis*, Pacific Grove, CA, 2005.

- [7] S. Tanaka, *Fault Diagnosis in Dynamic Systems: Theory and Application*. Prentice Hall International, 1989, ch. Diagnosability of Systems and Optimal Sensor Location, pp. 21–45.
- [8] S. J. Chang, F. DiCesare, and G. Goldbogen, "Evaluation of diagnosability of failure knowledge in manufacturing systems," in *The IEEE International Conference on Robotics and Automation*, 1990, pp. 696–701.
- [9] S. Chien, R. Doyle, and N. Rouquette, "Sensor placement for diagnosability in space-borne systems: A model-based reasoning approach," in *The 2nd International Workshop on Principles of Diagnosis*, Milan, Italy, 1991.
- [10] M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzi, "Diagnosability of discrete event system," *IEEE Transactions on Automatic Control*, vol. 40, no. 9, pp. 1555–1575, 1995.
- [11] P. Bouyer, F. Chevalier, and D. D'Souza, "Fault diagnosis using timed automata," in *Proc. of the 8th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'05)*, 2005, pp. 219–233.