# A Model-Based Framework for Stochastic Diagnosability

**Gregory Provan**
Computer Science Department,
University College Cork,
Cork, Ireland
email: g.provan@cs.ucc.ie

## Abstract

We propose a model-based framework to provide a clear semantics and well-developed algorithms for assessing the diagnosability of stochastic systems. This framework applies techniques originally developed for model-based diagnosis. Our framework enables us to (1) assess the diagnosability of a system, (2) analyze which faults are diagnosable, (3) identify sensor suites necessary for achieving pre-specified diagnosability levels.

## 1 Model-Based Diagnosability of Stochastic Systems

It is important to characterise the diagnosability of stochastic systems, as many important systems are inherently stochastic. Up until now, although there is significant literature on diagnosability of deterministic model-based systems (e.g., [3, 7]), there is no model-based framework for diagnosability of stochastic systems using a model-based diagnosis framework (e.g., as based on [12]). This article provides the first such framework.

We extend Reiter's consistency-based formulation of model-based diagnosis [12] to define diagnosability properties of stochastic systems using a model-based diagnosis framework. We adapt Reiter's model-based approach by extending model-based diagnosis principles for deterministic diagnosability [3, 2] to encode stochastic behaviour of the system. We have developed a theoretically well-founded framework and implemented algorithms to assess the diagnosability of stochastic systems. In this article we focus on static models, but all notions can be extended to cover temporal models. Given the model-based representation for finite-discrete-domain systems, we have analysed the theoretical properties of the models, providing information to:

- Assess the diagnosability of a stochastic system, using a model of the system;
- Rank the faults in terms of the expected loss of (or damage inflicted by) the fault;
- Identify the diagnosable faults via the suite of sensors defined in a system's model;
- Identify the sensors necessary for achieving pre-specified diagnosability levels.

We organize the remainder of the document as follows. Section 2 introduces our notation. Section 3 focuses on our definitions of stochastic diagnosability. Section 4 presents an example of a modeling framework using Bayesian networks. Section 5 shows how the systems modeled in our framework can be readily analysed to determine their diagnosability and detectability. Finally, Section 6 summarises our results.

## 2 Notation

This section introduces our notation for diagnosis systems. The standard consistency-based formulation of model-based diagnosis [12] characterises a model-based diagnosis (MBD) problem $\Psi$ using the triple $\langle \text{COMPS}, \text{SD}, \text{OBS} \rangle$:

- COMPS=$\{C_1, ..., C_m\}$ describes the operating modes of the set of $m$ components that comprise the system.

- SD, or system description, describes the function of the system. This model specifies two types of knowledge, denoted SD=$(\mathcal{S}, \mathcal{B})$: (1) system structure $\mathcal{S}$, i.e. the connections between the components, and (2) system behaviour $\mathcal{B}$, based on the behaviour of each component.
- OBS, the set of observations, denotes possible sensor measurements, which may be control inputs, outputs or intermediate variable-values.

In Reiter's theory, if we instantiate the $C_i \in$ COMPS$^* \subseteq$ COMPS to $\neg$ OK, and all other $C_i$ to OK, then we can formalise the notion of a diagnosis as the subset COMPS$^*$ of instantiated fault-mode variables such that SD $\cup$ OBS $\cup$ COMPS $\not\models \perp$ for an observation OBS.[1]

We extend this approach as follows. We characterise a system model using standard dynamical systems notation. We define a model in terms of a variable set $\mathcal{V}$, which consists of three disjoint subsets $\langle V, \Phi, \mathcal{Y} \rangle$, where:

- $V = \{V_1, ..., V_n\}$ is the set of state variables;
- $\Phi = \{\phi_1, ..., \phi_m\}$ is the set of fault variables (assumables) corresponding to COMPS;
- $\mathcal{Y} = \{Y_1, ...Y_s\}$ is the set of measured variables (observables) corresponding to OBS.

SD can now be defined as a set of constraints over the set $\mathcal{V}$ of variables, where COMPS$\cup$ OBS $\subseteq \mathcal{V}$, such that COMPS $\cap$ OBS $= \emptyset$. Component $i$ has associated mode-variable $C_i$, which takes on values such as $\{OK, Bad\}$, denoting that component $i$ is working normally or abnormally, respectively. We denote $C_i$ taking on value OK as $[C_i = OK]$. We assume in this article that a component may have multiple possible failure mode values, i.e., we do not assume only binary-valued failure-mode variables.

We will denote $\tilde{X}$ as the domain of $X$. We denote the set of all instantiations of a variable $X$ as $\mathbf{X}$. An observation $\vec{\mathcal{Y}}$ is an instantiation of a subset of $\mathcal{Y}$, i.e., $\vec{\mathcal{Y}} \subseteq \times_i \tilde{\mathcal{Y}}_i$.

Our model representation thus is $\Psi = (\mathcal{V}, \mathcal{S}, \mathcal{B})$, given a set $\mathcal{S} \cup \mathcal{B}$ of constraints over $\mathcal{V}$, and the input to a diagnosis problem is $(\Psi, \vec{\mathcal{Y}})$.

**Definition 1 (Candidate Diagnosis).** A candidate diagnosis $\Delta(\Psi, \vec{\mathcal{Y}})$ is an instantiation $\vartheta_{ij}$ of fault mode $\phi_i \in \Phi$, $i = 1, ..., m$, i.e., an assignment $\bigwedge_{i=1}^{m} [\phi_i = \vartheta_{ij}]$.

Given the collection of all possible multiple-fault diagnoses contained in the power-set $\mathcal{P}(\mathbf{\Phi})$, we define a *fault* $\Delta \in \mathcal{P}(\mathbf{\Phi})$ as a candidate diagnosis in which some $\phi_i \in \Phi$ takes on a value that is not OK.

We can define the notion of *ranked* deterministic diagnosis, using a ranking function $\varphi$.[2]

**Definition 2 (Ranking Function).** Given an observation $\vec{\mathcal{Y}}$, a *ranking function* is a function $\varphi$ that imposes an ordering on faults, e.g., $\varphi : \phi \rightarrow \mathbb{R}$ maps faults to real numbers.

We can use this ranking function to define the most likely diagnosis:

**Definition 3 (Most-likely Diagnosis).** Given an observation $\vec{\mathcal{Y}}$, a *most-likely diagnosis* is a fault $\Delta^*$ defined over $\phi$ such that $\Psi \cup \vec{\mathcal{Y}} \models \Delta^*$ and $\nexists \Delta$ such that $\Psi \cup \vec{\mathcal{Y}} \models \Delta$ and $\varphi(\Delta) < \varphi(\Delta^*)$.

We now define the notion of stochastic diagnosis. In a stochastic model, we define the model constraints $\mathcal{B}$ using a probability distribution over our system tuple $\langle V, \Phi, \mathcal{Y} \rangle$: $\mathcal{B} = Pr(V, \Phi, \mathcal{Y})$.

**Definition 4 (Stochastic Diagnosis).** A stochastic diagnosis $\Delta(\Psi, \vec{\mathcal{Y}})$ is an assignment of (non-zero) probabilities $Pr(\Delta|\vec{\mathcal{Y}})$ to $\Delta \in \mathcal{P}(\mathbf{\Phi})$ .

We extend Lucas's definition of diagnostic Bayesian inference [9] by assigning probability distributions not just to component failure modes (as Lucas does), but also to other system variables. Our definition of stochastic diagnosis is completely consonant with that of Lucas, who defines the presence of a fault in terms of a non-zero probability being assigned to a fault (or abnormality).

We can define a strong notion of diagnosis, in terms of the maximum a-posteriori explanation

[1] Reiter [12] uses a predicate logic in which an $AB$ predicate denotes abnormal behaviour; hence it defines $AB(c)$ are component $c$ being faulty. In this article we assume a propositional framework.

[2] Many ranking functions have been used in the model-based diagnosis literature, such as subset-minimality [6], probability [9], and qualitative probability [5]. We adopt a Bayesian ranking function that maps faults to [0,1].

(MAP), the most probable configuration of fault-mode variables given observation $\vec{\mathcal{Y}}$.

**Definition 5 (MAP Diagnosis).** Given observation $\vec{\mathcal{Y}}$, the most probable explanation, denoted $\Delta_{MAP}(\Delta, \vec{\mathcal{Y}})$, is given by $\overset{argmax}{\Delta}$ $Pr\{\Delta|\vec{\mathcal{Y}}\}$, and is an instantiation of assumable variables $\phi \in \Delta$ such that $Pr\{\Delta|\vec{\mathcal{Y}}\}$ is maximised.

## 3 Diagnosability

This section defines notions of fault diagnosability by extending the notions defined for deterministic systems [3, 7] to stochastic systems.

### 3.1 Deterministic Diagnosability

**Definition 6 (Sensor Signature).** *We define the sensor signature of a fault $\Delta$ as given by $\Psi \cup \Delta \models \vec{\mathcal{Y}}$.*

Loss of diagnosability occurs when two different faults trigger the same sensor signature.

**Definition 7 (Fault Diagnosability).** *A fault is diagnosable, given an appropriate observation $\vec{\mathcal{Y}}$ based on sensor suite $\Sigma$, if the fault can be uniquely isolated, i.e., we can compute a unique sensor signature $\vec{\mathcal{Y}}$. In other words, two faults $\Delta_i$ and $\Delta_j$, with respective signatures $\vec{\mathcal{Y}}_i$ and $\vec{\mathcal{Y}}_j$, are diagnosable under the sensor set $\mathcal{Y}$ only if $\vec{\mathcal{Y}}_i \neq \bar{\vec{\mathcal{Y}}}_j$.*

We can extend this notion to *system diagnosability* as follows:

**Definition 8 (System Diagnosability).** *A system is diagnosable, given a sensor set $\mathcal{Y}$, if and only if (a) every fault can be uniquely isolated given an appropriate observation $\vec{\mathcal{Y}}$ based on $\mathcal{Y}$, and (b) for any observation $\vec{\mathcal{Y}}$ there exists a unique most-likely fault.*

We can formalise this as follows:

**Lemma 1 (Diagnosability).** *For a system $\Psi$ with sensors $\mathcal{Y}$ and fault variables $\phi$, a fault $\Delta$ is diagnosable via observation $\vec{\mathcal{Y}}$ iff*

1. *$\exists$ a unique $\Delta$ such that $[\Psi \cup \vec{\mathcal{Y}} \models \Delta] \wedge [\Psi \cup \Delta \models \vec{\mathcal{Y}}]$. A fault is undiagnosable if $[\Psi \cup \vec{\mathcal{Y}} \not\models \Delta] \wedge [\Psi \cup \Delta \models \vec{\mathcal{Y}}]$.*
2. *for every value of an assumable such that $[\phi \neq \text{nominal}]$, we have some $\vec{\mathcal{Y}}$ such that $\Psi \cup \vec{\mathcal{Y}} \models \phi$.*

The proposed definitions of diagnosability differ from the diagnosability definitions of [3, 7] in some important ways. We use our consistency-based definition of fault as the basis for the definition of (and method of computing) diagnosability, whereas [3, 7] use different definitions; for example, [7] uses a set-theoretic definition. In addition, we adopt a decision-theoretic approach, such that we can define loss-minimal notions of fault and diagnosability; neither [3, 7] define loss functions to enable such notions.

### 3.2 Stochastic Diagnosability

We need to frame our definitions of stochastic diagnosability in terms of avoiding situations of undiagnosability, since diagnosability is now dealing with continuous spaces, as opposed to the deterministic model, for which sharp distinctions on discrete spaces can be made.

We first define when a fault and a system are undiagnosable.

**Definition 9 (Stochastic Fault Undiagnosability).** *A fault $\Delta_i$ with associated sensor measurement $\vec{\mathcal{Y}}_j$, is undiagnosable (undetectable) if $Pr(\Delta_i|\vec{\mathcal{Y}}_j) = 0 \quad \forall \vec{\mathcal{Y}}_j \in \vec{\mathcal{Y}}$.*

**Definition 10 (Stochastic System Undiagnosability).** *A system $model$ is undiagnosable if any possible fault is undiagnosable, i.e, if for any $\Delta_i \in \mathcal{P}(\mathbf{\Phi})$ with associated sensor measurement $\vec{\mathcal{Y}}_j$, $Pr(\Delta_i|\vec{\mathcal{Y}}_j) = 0 \quad \forall \vec{\mathcal{Y}}_j \in \vec{\mathcal{Y}}$.*

We now define two classes of stochastic diagnosability, weak and strong. These classes differ in terms of the strength of evidence required to establish one class or another.

**Definition 11 (Weak Stochastic Diagnosability).** *A system $model$ is weakly diagnosable if there is no $\Delta_i \in \mathcal{P}(\mathbf{\Phi})$ that is undiagnosable.*

We now shift our attention to fault diagnosability for stochastic systems. The notion of diagnosability is quite clear-cut for deterministic systems, and for stochastic systems requires a domain-dependent $(0, 1)$ threshold, as in Definition 12, or an MAP-based notion, as in Definition 13.

**Definition 12 (Stochastic Diagnosability).** *For a stochastic system $\Psi$ with sensors $\mathcal{Y}$ and fault variables $\phi$, a fault $\Delta$ is diagnosable via obser-*

*vation $\vec{\mathcal{Y}}$ iff $Pr\{\Delta|\Psi, \vec{\mathcal{Y}}\} > \delta$ for some (domain-dependent) threshold $\tau \in (0, 1)$, and is undiagnosable otherwise.*

**Definition 13 (Strong Stochastic Diagnosability).** *For a stochastic system $\Psi$ with sensors $\mathcal{Y}$ and fault variables $\phi$, a fault $\Delta$ is strongly diagnosable via observation $\vec{\mathcal{Y}}$ if $\Delta$ is the MAP diagnosis, i.e., $\Delta = \overset{argmax}{\Delta} Pr\{\Delta|\vec{\mathcal{Y}}\}$.*

### 3.3 Minimal-Loss Diagnosability

This section extends our definitions of stochastic diagnosability (which incorporates only uncertainty) to a decision-theoretic framework that incorporates both uncertainty and the loss associated with a system state. For example, we can assign a loss measure associated with a device functioning normally or if it has a fault. This notion is standard in real-world applications, such as a car manufacturing plant, which measures the downtime caused by a failure in thousands of euro per hour.

Our objective is to compute the loss associated with a fault, using a utility (or loss) function $\mu$, which specifies the loss (utility) associated with an action $\mathcal{A}$, given a world state, i.e., $\mu : \mathcal{A} \times 2^{\phi} \rightarrow \mathbb{R}$..[3] Given a likelihood ranking over faults, we can compute the expected loss associated with faults, i.e., a ranking of which faults cause the most damage to the diagnosis system, using a utility function of the form $\mu : \Delta \rightarrow \mathbb{R}$. Given a set of utilities $\{\mu_1, ..., \mu_r\}$ and likelihoods $\{\varphi_1, ..., \varphi_r\}$ associated with diagnoses $\{\Delta_1, ..., \Delta_r\}$, we can compute the expected utility as $E[\mu(\Delta)] = \sum_{i=1}^{r} \varphi_i * \mu_i(\Delta)$.

We can use this expected utility to define a minimum-loss fault:

**Definition 14.** Given an observation $\vec{\mathcal{Y}}$, a *minimum-loss fault* $\Delta^*$ is a fault where $\Psi \cup \vec{\mathcal{Y}} \models \Delta^*$ and $\nexists \Delta$ such that $\Psi \cup \vec{\mathcal{Y}} \models \Delta$ with $\mu(\Delta) < \mu(\Delta^*)$.

## 4 Example: Bayesian Network Diagnosability

The modeling framework that we extend is a Bayesian Network (BN) [10]. Our underlying model representation is a directed acyclic graph (DAG) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of vertices $\mathcal{V}$ and edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. We denote a directed edge from $V_i$ to $V_j$ by $(V_i, V_j)$. The edges show independence relations, such that the absence of an edge from $V_i$ to $V_j$ denotes that $V_j$ is independent of $V_i$.

**Definition 15 (Bayesian Network).** A BN is a tuple $(\mathcal{G}, \zeta)$, where $\mathcal{G}$ is a DAG, and $\zeta$ is a set of *probability distributions* constructed from vertices in $\mathcal{G}$ based on the topological structure of $\mathcal{G}$. $\zeta$ satisfies $Pr\{\mathcal{V}\} = \prod_{i=1}^{n} Pr\{V_i|pa(V_i)\}$, where $pa(V_i)$ are the parents of $V_i$ in $\mathcal{G}$.

We can use BNs to model MBD problems by:

- assigning each MBD variable in COMPS to a special BN node representing a failure-mode variable $\phi$;
- assigning each non-mode MBD variable $V_i \in \mathcal{V} \setminus COMPS$ to a BN node in $\mathcal{V} \setminus \phi$;
- defining the MBD system structure $\mathcal{S}$ using the BN graphical structure, and the MBD system behaviour $\mathcal{B}$ using the CPTs of the BN.

A BN allows you to compute the posterior distribution for any variable, given an observation $\vec{\mathcal{Y}}$.

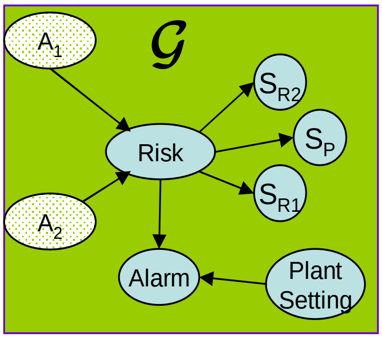


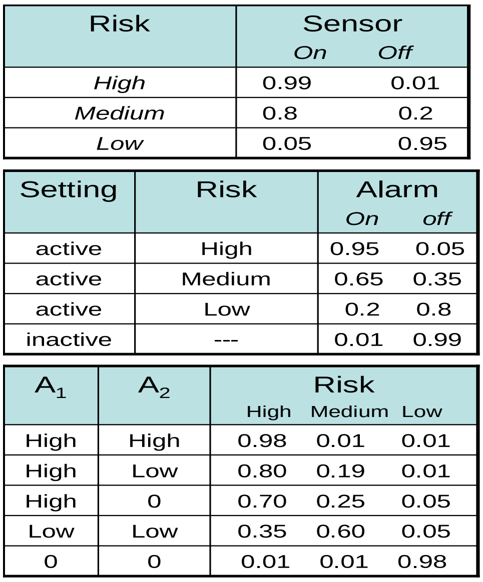
Figure 1: Causal graph for alarm activation based on fault status in a power plant.

We can apply our BN framework to an example as follows. Figure 1 depicts a causal graph for diagnosing faults in a simplified model of a nuclear power plant, which consists of a cooling tower and two reactors, with fault-detection sensors $S_P, S_{R1}, S_{R2}$ respectively. Faults in the system will sound an alarm, depending on the plant operational setting *Setting* (either *active* or *inactive*), plus the risk level for a fault, denoted by

[3] In this article we assume that the diagnosis system just identifies faults and no other actions are taken, i.e., $\mathcal{A} = \emptyset$.

variable *Risk*, with values $\{$*high, medium, low*$\}$. The risk level is itself determined by the values of two fault-mode variables, $A_1$, $A_2$, denoting the presence of a fault at levels $\{$*high,low,zero*$\}$ in reactors $R_1$ and $R_2$ respectively. The sensors indicate an alarm status, which takes on values $\{$*present, absent*$\}$.

Figure 2 shows a subset of the Conditional Probability Tables for Bayesian network. Given observation $\vec{\mathcal{Y}} = \{[S_{W_1}$=*off*], [$S_{W_2}$=*alarm*], [$S_D$ =*alarm*] and [alarm=*on*]$\}$, we can compute the most likely (MAP) single faults $\Delta$ as [$A^1_{int}$=*low*] and [$A^2_{int}$=*low*], each with probability 0.420. Given this suite of sensors, it turns out that the system is both strongly and weakly diagnosable, with a threshold of 0.3. However, if we restrict our observables to just the alarm (and not the individual sensors as well), then the system needs a significantly lower threshold, of the order of 0.15. In this case only weak diagnosability pertains. This indicates that the system is quite dependent on the suite of sensors.



| Risk | Sensor | |
|---|---|---|
| | *On* | *Off* |
| *High* | 0.99 | 0.01 |
| *Medium* | 0.8 | 0.2 |
| *Low* | 0.05 | 0.95 |

| Setting | Risk | Alarm | |
|---|---|---|---|
| | | *On* | *off* |
| active | High | 0.95 | 0.05 |
| active | Medium | 0.65 | 0.35 |
| active | Low | 0.2 | 0.8 |
| inactive | --- | 0.01 | 0.99 |

| $A_1$ | $A_2$ | Risk | | |
|---|---|---|---|---|
| | | High | Medium | Low |
| High | High | 0.98 | 0.01 | 0.01 |
| High | Low | 0.80 | 0.19 | 0.01 |
| High | 0 | 0.70 | 0.25 | 0.05 |
| Low | Low | 0.35 | 0.60 | 0.05 |
| 0 | 0 | 0.01 | 0.01 | 0.98 |

Figure 2: Partial description of Conditional Probability Tables for Bayesian network. The dash in the second table indicates that the probabilities for alarm, given [*Setting*=inactive] are independent of the value of *Risk Status*.

## 5 Diagnosability Assessment

This section describes an algorithm for computing system diagnosability, and how we can analyse the resulting levels of system diagnosability.

### 5.1 Diagnosability Measures

Since it may not be possible to guarantee absolute diagnosability for a real-world system, one key parameter of interest is the percentage of faults that are actually diagnosable. To compute this, we need the following notions.

We use the indicator variable $\delta_i$ to indicate diagnosability of the $i^{th}$ fault. Given a probability threshold $\tau$, we denote a threshold-based $\delta_i$ as follows:

$$\delta_i = \begin{cases} 1 & \text{if } Pr(\Delta_i|\vec{\mathcal{Y}}) > \tau, \\ 0 & \text{otherwise} \end{cases}$$

Note that we can use an alternative, MAP-based, definition of $\delta_i$, namely

$$\delta_i = \begin{cases} 1 & \text{if } \Delta_i = \overset{argmax}{\Delta} \; Pr(\Delta|\vec{\mathcal{Y}}), \\ 0 & \text{otherwise} \end{cases}$$

The *diagnosability measure* of a model $\Psi$ is the fraction of all faults that can be identified:

**Definition 16.** The *diagnosability measure* of $\Psi$ is given by

$$\kappa(\Psi, \mathcal{Y}) = \frac{\sum_i \delta_i : i \in \mathcal{P}(\mathbf{\Phi}) - 1}{\mathcal{P}(\mathbf{\Phi}) - 1} \qquad (1)$$

We can also use loss-weighted diagnosability measures, but we omit their description here for ease of exposition.

### 5.2 System Diagnosability Algorithm

We have developed algorithms for estimating a system's diagnosability measure by generalising the diagnosability approach described in [11]. Since computing diagnosability measures is computationally taxing for complex system models, we employ approximation techniques, e.g., focusing search on the highest-loss faults.

We now describe an algorithm for computing system diagnosability.[4] We adopt an approach dif-

[4] Note that we can map our representation into one where it is possible to use a variety of verification tools, such as SMV/NUSMV [1], or into a timed (stochastic) automaton [4].

ferent from the techniques for computing diagnosability [13, 2], which use formal methods, together with the list of possible faults, to determine if faults can be identified uniquely.

We first simulate the sensor-outputs predicted given the presence of an fault-set, and then use the sensor-outputs to try to diagnose the presence of the true faults.

We assess which of the vulnerable system states will result in alarms being raised, which aids in computing a diagnosability measure. Our diagnosability analysis procedure consists of three main steps.

1. First, for each fault $\Delta$ in a set of faults to test, $\mathcal{P}(\mathbf{\Phi})$, simulate the associated test- (or sensor-vector) $T$. This produces a set of pairs of faults and test-vectors, $\{\langle\Delta, T\rangle\}$.[5]
2. Next, for each test-vector, compute the fault-list associated with that vector, generating the tuple $\{\langle\Delta, T, \alpha\rangle\}$.
3. Finally, compute the diagnosability statistics given the tuples.

To facilitate this procedure, we have two models: a simulation model $\Psi_S$, and a diagnosis model $\Psi_D$. Each model is optimized for its particular task, either simulation or fault-detection. Our key concern is to specify the difference between the real fault-state $\Delta$ and the computed fault-state (as represented by the fault $\alpha$); we average this value over a large suite of real faults $\Delta$. If $\alpha = \Delta$, then we have perfect fault detection. In many cases there is an ambiguity group (set of disjunctive fault-states $\alpha = \alpha_1 \vee \cdots \vee \alpha_l$), so we must compute the degree of diagnosability that we have. If $\alpha = \emptyset$, then that fault is undiagnosable.

### 5.3 Sensor Selection and Placement

A key capability of any diagnosis system is the ability to isolate faults when they occur. This capability is a function of the design of the system, and in particular of the sensors that the system has for state-measurement.

The general sensor selection task is a difficult one, and can be formulated as follows [8]. Given a set $\mathcal{Y} = \{Y_1, ..., Y_s\}$ of sensors (together with their locations), we need to choose the set of sensors to optimise a function $\Gamma$. This task requires trading off the utility associated with optimising $\Gamma$ with the cost of the sensors and their associated measurements.

We assume a utility function $\eta : 2^{\mathcal{Y}} \rightarrow \mathbb{R}$, and a cost function $\chi : 2^{\mathcal{Y}} \rightarrow \mathbb{R}$, which returns the cost of measurements from each $\mathcal{Y}' \subseteq \mathcal{Y}$. The sensor selection problem then is to choose the subset $\mathcal{Y}' \subseteq \mathcal{Y}$ that maximises utility and minimises cost. We can define a generic function for this task that computes the highest-utililty subset of sensors for a cost limit of $\beta$.

$$SS(\beta) \;=\; \overset{argmax}{\mathcal{Y}' \subseteq \mathcal{Y}, \chi(\mathcal{Y}') \leq \beta} \, \eta(\mathcal{Y}').$$

In our case, we can define the diagnosability problem as a sensor selection problem where our utility function is the diagnosability measure $\kappa(\Psi, \mathcal{Y})$, and the cost function assumes each sensor has identical unit cost and measurements are free.

### 5.4 Achieving Target Diagnosability Levels

Real-world systems typically entail trading off the cost of sensors for system observability. In the case of fault detection, we can use our diagnosability assessment techniques to analyse appropriate tradeoffs of sensors for fault diagnosability. In other words, we can compute the suite of sensors necessary to achieve a [0,1] fault diagnosability measure. To accomplish this, we need to extend our definition of system model to include the sensor suite $\Sigma$, which specifies the number and topological location of the sensors.

We can define the fixed-target diagnosability (FTD) problem as a sensor selection problem where we impose a targeted diagnosability level of $\tau$, our utility function is the diagnosability measure $\kappa(\Psi, \mathcal{Y})$, and the cost function assumes each sensor has identical unit cost and measurements are free. We can formulate this problem as follows:

[5] This step is analogous to the Diagnosability Assessment step, as it is simulating the sensor settings that should be recorded if a fault were compromised. Note that if we have insufficient sensors, then a fault could be compromised without any alarms being raised. $F$ is the true Diagnosability-state, and the sensor vector $T$ allows us to identify the recorded state.

$$FTD(\beta) \;=\; \overset{argmin}{\mathcal{Y}\subseteq\mathcal{Y},\kappa(\Psi,\mathcal{Y}')\geq\tau}\; \chi(\mathcal{Y}').$$

Our analysis is related to the use of analytical redundancy relations (ARR) for diagnosability [14]. The use of ARRs is based on the notion of fault signatures, which define the faults in a fault signature matrix. In contrast, we examine a sensor-signature matrix, in which the sensor signatures are analysed. In situations where there are fewer sensor signatures than fault signatures (which we argue is most likely the case given that most situations have limited sensors), this approach will be more efficient than the ARR-based approach.

We now analyse the case of having $k$ binary sensors.[6] We can define a sensor signature $\mathcal{S}_i$ of an fault $\phi_i$ as the vector of sensor values under $\phi$.

**Definition 17 (Deterministic Fault Sensor Signature).** *Given a set $\mathcal{Y}$ of $s$ sensors, the fault sensor signature of fault $\Delta_j$ is a binary vector $\{\sigma_{1j}, \sigma_{2j}, ..., \sigma_{sj}\}$ in which*

$$\sigma_{ij} = \begin{cases} 1 & \textit{if } \Delta_j \textit{ causes an anomalous reading at } \vec{\mathcal{Y}}_i, \\ 0 & \textit{if } \Delta_j \textit{ causes a nominal reading at } \vec{\mathcal{Y}}_i. \end{cases}$$

We generalise this deterministic signature to a stochastic signature as follows.

**Definition 18 (Stochastic Fault Sensor Signature).** *Given a set $\mathcal{Y}$ of $s$ sensors and threshold $\tau \in [0,1]$, the stochastic fault sensor signature of fault $\Delta_j$ is a vector $\{\sigma_{1j}, \sigma_{2j}, ..., \sigma_{sj}\}$ in which*

$$\sigma_{ij} = \begin{cases} 1 & \textit{if } Pr(\Delta_j|\mathcal{Y}_i = \textit{on}) > \tau, \\ 0 & \textit{if } Pr(\Delta_j|\mathcal{Y}_i = \textit{nominal}) \geq (1-\tau). \end{cases}$$

A fault sensor matrix $M(\Psi, \mathcal{Y})$ is a $q \times r$ matrix in which row $i,\ i = 1, ..., r$, consists of the fault sensor signature for fault $\Delta_i$. We use this sensor data to define the discriminability of a suite of sensors. We can discriminate two faults using Definition 7, i.e., faults $\Delta_i$ and $\Delta_j$ are discriminable under sensors $\mathcal{Y}$ if $\vec{\mathcal{Y}}_i \neq \vec{\mathcal{Y}}_j$. We can extend this notion to system-wide discriminability, i.e., where all sensors are pairwise discriminable. We can identify the required sensors for discriminability as follows.

**Lemma 2.** *If we have $k$ binary sensors, then the maximum number of descriminable faults is $2^k - 1$.*

**Corollary 1.** *The minimum number of binary-valued sensors required to discriminate $m$ faults is $\lceil log_2(m+1)\rceil$.*

Given a potential suite of $N$ possible sensors, we can identify a subset of cardinality-minimal sensors that can achieve a target diagnosability measure as follows.[7] If $M(\Psi, \mathcal{Y})$ is a $s \times r$ matrix of rank $\geq \lceil log_2(r+1)\rceil$ then all faults can be detected, i.e., the diagnosability measure is 1.[8] If the rank is smaller, then we must compute the diagnosability measure from $M(\Psi, \mathcal{Y})$ using Equation 1.

For fault detectabilty analysis, we have adapted a system (based on [11]) that allows a user to compare the diagnosability measures of a set of models, e.g., $\Psi(\mathcal{Y}_1), ..., \Psi(\mathcal{Y}_m)$, that differ only in the relative position, number and type of sensors. We generate $M(\Psi, \mathcal{Y})$ and use this to see if $\Delta(\Psi) > \tau$, where $\tau$ is a target diagnosability level.

### 5.5 Example

We have two fault variables, $A_1$ and $A_2$, each with 3 possible values {*high, low, nominal*}, giving $2^3 - 1$ faults, which will require at least 4 binary sensors to achieve fault discriminability (using Corollary 1). Table 1 depicts the sensor signatures for the single-fault scenarios.

Consider the case where we limit our analysis to faults occurring only one at a time, e.g., we can have a single fault $[A_1 = high]$. Given the current suite of sensors $\{S_{R_1},\ S_{R_2},\ S_P\}$, Table 1 shows the single-fault sensor signatures, from which we can discriminate only $[A_1 = low]$ and $[A_2 = low]$, since the sensor signatures for the cases for both *high* and *low* fault levels are the same for $A_1$ and $A_2$. Hence, we can compute the diagnosability measure $\Delta(\Psi, \mathcal{Y})$ as $\frac{1}{3}$, since we can discriminate only 2 out of 6 total faults.

[6] We can extend this case to $q$-ary sensors in a direct manner.

[7] We can use other criteria for sensor selection, such as cost, by identifying an objective function to optimise during this sensor minimisation process.

[8] See [14] for justification of this approach.

| $A_1$ | $S_{R_1}$ | $S_{R_2}$ | $S_P$ | $A_2$ | $S_{R_1}$ | $S_{R_2}$ | $S_P$ |
|---|---|---|---|---|---|---|---|
| high | 1 | 1 | 1 | high | 1 | 1 | 1 |
| high | 1 | 1 | 0 | high | 1 | 1 | 0 |
| high | 1 | 0 | 1 | high | 1 | 0 | 1 |
| low | 1 | 0 | 0 | low | 0 | 1 | 1 |
| low | 0 | 1 | 0 | low | 0 | 1 | 1 |
| low | 0 | 0 | 1 | low | 0 | 1 | 1 |
| nominal | 0 | 0 | 0 | nominal | 0 | 0 | 0 |

Table 1: Fault sensor matrix, with sensor signatures defined for the two fault types, $A_1$ and $A_2$.

## 6 Concluding Remarks

This article has presented a general framework for computing system diagnosability for model-based systems. This approach provides several important advantages over existing approaches. The model-based approach enables a user to define a range of levels of detail in a model and yet still retain verification and diagnosability properties, facilitating trading off the cost of model construction with the quality of fault isolation and repair.

## References

[1] A. Cimatti, E. M. Clarke, F. Giunchiglia, and M. Roveri. NUSMV: A new symbolic model verifier. In *Computer Aided Verification*, pages 495–499, 1999.

[2] A. Cimatti, C. Pecheur, and R. Cavada. Formal verification of diagnosability via symbolic model checking. In *Proceedings of IJCAI'03*, Acapulco, Mexico, 2003.

[3] L. Console, C. Picardi, and M. Ribaudo. Diagnosis and Diagnosability Analysis Using PEPA. In *14th European Conference on Artificial Intelligence (ECAI)*, pages 131–135, Berlin, Germany, 2000.

[4] R. Corin, S. Etalle, P. H. Hartel, and A. Mader. Timed analysis of security protocols. *J. Computer Security*, to appear, 2006.

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

[6] J. de Kleer, A. K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56(2-3):197–222, 1992.

[7] O. Dressler and P. Struss. A toolbox for integrating model-based diagnosability analysis and automated generation of diagnostics. In *Proceedings of DX'03*, pages 99–104, Washington, D.C., USA, 2003.

[8] V. Isler and R. Bajcsy. The sensor selection problem for bounded uncertainty sensing models. *IEEE Tran. Automation Science and Engineering*, to appear, 2006.

[9] P. Lucas. Bayesian model-based diagnosis. *International Journal of Approximate Reasoning*, 27:99–119, 2001.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[11] G. Provan. Diagnosability analysis for distributed systems. In *Proc. IEEE Aerospace Conference*, Big Sky, USA, March 2002.

[12] R. Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 32:57–96, 1987.

[13] M. Sampath, R. Sangupta, S. Lafortune, K. Sinnamohideen, and D. Teneketzis. Diagnosibility of discrete event systems. *IEEE Transactions on Automatic Control*, 40(9):1555–1575, Sept. 1995.

[14] L. Trave-Massuyes, T. Escobe, and R. Milne. Model-based diagnosability and sensor placement application to a frame 6 gasturbine subsystem. In *Proc. of 12th Intl. Workshop on Principles of Diagnosis*, pages 205–212, Sansicario, Via Lattea, Italy, March 2001.