Reducing Model-Based Diagnosis to Knowledge Base Debugging

Patrick Rodler, Konstantin Schekotihin

Alpen-Adria Universität Klagenfurt
# Contents

1. Overview
2. Model-based Diagnosis (MBD)
3. Knowledge Base Debugging (KBD)
4. Reduction (MBD → KBD)
5. Example
6. Conclusions
Model-based Diagnosis (MBD) is a principled approach to fault localization in any type of system that can be described in a formal structured way.

Knowledge Base Debugging (KBD) draws on concepts from MBD to find faults in a monotonic knowledge base.

We show that KBD is a generalization of MBD in that any MBD problem can be reduced to a KBD problem, and solutions of the MBD problem can be directly extracted from solutions of the KBD problem.

The sequential MBD problem is a special case of the sequential KBD problem in that the latter allows a user to provide more types of measurements and specify additional requirements (beyond consistency).

Consequently: KBD approaches can be applied to all systems amenable to MBD.
Definition 1 (System)

A *system* is a tuple \((SD, COMPS)\) where \(SD\), the system description, is a set of first-order sentences, and \(COMPS\), the system components, is a finite set of constants \(c_1, \ldots, c_n\).

Let \(SD_{beh} := \{ \neg \text{AB}(c) \rightarrow \text{beh}(c) \mid c \in COMPS\}\) where \(\text{beh}(c)\) denotes the first-order sentence describing the expected behavior of \(c \in COMPS\). General axioms describing the system domain or descriptions of the interplay between the system components are comprised by \(SD_{gen}\). So, \(SD = SD_{beh} \cup SD_{gen}\).

Definition 2 (MBD-DPI)

Let \(OBS\) (observations) be a finite set of first-order sentences, \(MEAS\) (measurements) be a finite set including finite sets \(m_i\) of first-order sentences, and \((SD, COMPS)\) be a system. Then \((SD, COMPS, OBS, MEAS)\) is an *MBD diagnosis problem instance (MBD-DPI)*.
Definition 3 \((SD^*[\Delta])\)

Let \(DPI := (SD, COMPS, OBS, MEAS)\) be an MBD-DPI and \(U_{MEAS}\) denote the union of all \(m \in MEAS\). Then \(SD^*[\Delta] := SD \cup \{AB(c) \mid c \in \Delta\} \cup \{\neg AB(c) \mid c \in COMPS \setminus \Delta\} \cup OBS \cup U_{MEAS}\) for \(\Delta \subseteq COMPS\) denotes the behavior description of the system \((SD, COMPS)\)

- under the current state of knowledge given by the \(DPI\) in terms of OBS and MEAS, and
- under the assumption that all components in \(\Delta \subseteq COMPS\) are faulty and all components in \(COMPS \setminus \Delta\) are healthy.
Definition 4 (MBD-Diagnosis)

Let $DPI := (SD, COMPS, OBS, MEAS)$ be an MBD-DPI. Then $\Delta \subseteq COMPS$ is an **MBD-diagnosis for DPI** iff $SD^*[\Delta]$ is consistent (i.e. $\Delta$ explains OBS and MEAS). An MBD-diagnosis $\Delta$ for $DPI$ is called **minimal** iff there is no MBD-diagnosis $\Delta'$ for $DPI$ such that $\Delta' \subset \Delta$.

Problem 1 (Sequential MBD)

**Given:** An MBD-DPI $DPI := (SD, COMPS, OBS, MEAS)$ and a diagnostic goal $G$. **Find:** $MEAS_{new} \supseteq \emptyset$ and $\Delta$, where $MEAS_{new}$ is a set of new measurements such that $\Delta$ is a minimal MBD-diagnosis for the MBD-DPI $DPI_{new} := (SD, COMPS, OBS, MEAS \cup MEAS_{new})$ and $\Delta$ satisfies $G$.

**Remark:** Examples for diagnostic goals $G$ are the presence of one highly probable or just a single remaining (minimal) diagnosis.
Definition 5 (KBD-DPI)

Let

- $\mathcal{K}$ be a KB,
- $P, N$ be sets including sets of sentences,
- $R \supseteq \{ \text{consistency} \}$ be a set of (logical) requirements,
- $\mathcal{B}$ be a KB such that $\mathcal{K} \cap \mathcal{B} = \emptyset$ and $\mathcal{B}$ satisfies all requirements $r \in R$,
- the cardinality of all sets $\mathcal{K}, \mathcal{B}, P, N$ be finite, and
- all sets $\mathcal{K}, \mathcal{B}, P, N$ be formulated over some monotonic logic.

Then we call the tuple $\langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ a KBD diagnosis problem instance (KBD-DPI).
Definition 6 (Solution KB)

Let $DPI := \langle K, B, P, N \rangle_R$ be a KBD-DPI. Then a KB $K^*$ is called solution KB w.r.t. $DPI$ iff all the following conditions hold:

\[
\forall r \in R : \quad K^* \cup B \text{ fulfills } r \\
\forall p \in P : \quad K^* \cup B \models p \\
\forall n \in N : \quad K^* \cup B \not\models n.
\]

A solution KB $K^*$ w.r.t. $DPI$ is called maximal iff there is no solution KB $K'$ w.r.t. $DPI$ such that $K' \cap K \supset K^* \cap K$ (i.e. $K^*$ has a set-maximal intersection with $K$ among all solution KBs).
KBD-Diagnosis and KBD-Conflict

Definition 7 (KBD-Diagnosis)

Let $\text{DPI} := \langle K, B, P, N \rangle_R$ be a KBD-DPI. A set of sentences $D \subseteq K$ is called a $\text{KBD-diagnosis}$ w.r.t. DPI iff $(K \setminus D) \cup U_P$ is a solution KB w.r.t. DPI (i.e. $K^* := (K \setminus D) \cup U_P$ satisfies (1) – (3)). A KBD-diagnosis $D$ w.r.t. DPI is $\text{minimal}$ iff there is no $D' \subset D$ such that $D'$ is a KBD-diagnosis w.r.t. DPI.

Definition 8 (KBD-Conflict)

Let $\text{DPI} := \langle K, B, P, N \rangle_R$ be a KBD-DPI. A set of formulas $C \subseteq K$ is called a $\text{KBD-conflict}$ w.r.t. DPI iff $C \cup U_P$ is not a solution KB w.r.t. DPI (i.e. $K^* := C \cup U_P$ violates at least one of (1) – (3)). A KBD-conflict $C$ w.r.t. DPI is $\text{minimal}$ iff there is no $C' \subset C$ such that $C'$ is a KBD-conflict w.r.t. DPI.
In general, the (maximal) solution KB resulting from the deletion of one and the same set $\mathcal{D}$ from $\mathcal{K}$ is not unique. Since

- $P$ does not justify the inclusion of sentences (semantically) different from $U_P$, and
- only one solution KB is sought

we define:

**Definition 9 (Canonical Solution KB)**

$$(\mathcal{K} \setminus \mathcal{D}) \cup U_P$$

is the *canonical solution KB for $\mathcal{D}$ w.r.t. DPI* iff

$$(\mathcal{K} \setminus \mathcal{D}) \cup U_P$$

is a solution KB w.r.t. DPI.
The relationship between maximal canonical solution KBs and minimal KBD-diagnoses w.r.t. a DPI is as follows (cf. [Rodler, 2015]):

### Property 1

Let $DPI$ be a KBD-DPI. Then the set of all maximal canonical solution KBs w.r.t. $DPI$ is given by

$$\{(\mathcal{K} \setminus D) \cup U_P \mid D \text{ is a minimal KBD-diagnosis w.r.t. } DPI\}$$

Therefore, KBD methods focus on the computation of minimal KBD-diagnoses in order to find all maximal canonical solution KBs.
The relationship between the notions *KBD-diagnosis*, *solution KB* and *KBD-conflict* is as follows:

**Property 2**

Let $\mathcal{D} \subseteq \mathcal{K}$. Then the following statements are equivalent:

1. $\mathcal{D}$ is a KBD-diagnosis w.r.t. $\langle \mathcal{K}, \mathcal{B}, \mathcal{P}, \mathcal{N} \rangle_R$.
2. $(\mathcal{K} \setminus \mathcal{D}) \cup \mathcal{U}_P$ is a solution KB w.r.t. $\langle \mathcal{K}, \mathcal{B}, \mathcal{P}, \mathcal{N} \rangle_R$.
3. $(\mathcal{K} \setminus \mathcal{D})$ is not a KBD-conflict w.r.t. $\langle \mathcal{K}, \mathcal{B}, \mathcal{P}, \mathcal{N} \rangle_R$. 
Sequential KB Debugging

The sequential KBD problem which seeks a set of test cases in order to achieve a diagnostic goal $G$ is defined as follows:

**Problem 2 (Sequential KBD)**

**Given:** A KBD-DPI $DPI := \langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ and a diagnostic goal $G$.

**Find:** $P_{new}, N_{new} \supseteq \emptyset$ and $\mathcal{D}$, where $P_{new}, N_{new}$ are sets of positive and negative test cases, respectively, such that $\mathcal{D}$ is a minimal KBD-diagnosis w.r.t. $DPI_{new} := \langle \mathcal{K}, \mathcal{B}, P \cup P_{new}, N \cup N_{new} \rangle_R$ and $\mathcal{D}$ satisfies $G$. 
Reducing MBD to KBD

Any MBD-DPI can be modeled as a KBD-DPI, and the solutions of the latter directly yield the solutions of the former:

**Theorem 1 (Reduction of MBD to KBD)**

Let $mDPI := (SD, COMPS, OBS, MEAS)$ be an MBD-DPI where $COMPS = \{c_1, \ldots, c_n\}$. Then:

- $mDPI$ can be formulated as a KBD-DPI $kDPI$ such that there is a bijective correspondence between KBD-diagnoses for $kDPI$ and MBD-diagnoses for $mDPI$.
- All MBD-diagnoses for $mDPI$ can be computed from the KBD-diagnoses for $kDPI$. 
Reducing MBD to KBD

Proof:

We first show how \textit{mDPI} can be formulated as a KBD-DPI \textit{kDPI}. To this end, we specify how \textit{kDPI} = \langle \mathcal{K}, \mathcal{B}, P, N \rangle_R can be written in terms of the components of \textit{mDPI} = (\text{SD}_{beh} \cup \text{SD}_{gen}, \text{COMPS}, \text{OBS}, \text{MEAS}):

\[
\begin{align*}
\mathcal{K} &= \{ \alpha_i \mid \alpha_i := beh(c_i), c_i \in \text{COMPS} \} \tag{4} \\
\mathcal{B} &= \text{OBS} \cup \text{SD}_{gen} \tag{5} \\
P &= \text{MEAS} \tag{6} \\
N &= \emptyset \tag{7} \\
R &= \{ \text{consistency} \} \tag{8}
\end{align*}
\]
Reducing MBD to KBD

Proof (cont’d):

That is, $K$ captures $SD_{beh} \cup \{\neg AB(c_i) \mid c_i \in COMPS\}$, i.e. the nominal behavioral descriptions of all system components. By Def. 7, $D \subseteq K$ is a KBD-diagnosis for $kDPI$ iff both

$$ (K \setminus D) \cup B \cup UP \text{ meets all } r \in R \text{ (i.e. is consistent)} \quad (9) $$

and

$$ (K \setminus D) \cup B \cup UP \not= n \text{ for all } n \in N \quad (10) $$

hold.
## Proof (cont’d):

Let now $D$ be an arbitrary KBD-diagnosis for $kDPI$ such that $D = \{\alpha_i \mid i \in I\}$ for the index set $I \subseteq \{1, \ldots, n\}$.

Using (4) – (8) above, condition (9) for $D$ is equivalent to the consistency of

$$SD_{beh} \cup \{AB(c_i) \mid i \in I\} \cup \{\neg AB(c_i) \mid i \in \{1, \ldots, n\} \setminus I\} \cup OBS \cup SD_{gen} \cup U_{MEAS}$$

which in turn yields that

$$SD \cup \{AB(c_i) \mid c_i \in \Delta\}$$
$$\cup \{\neg AB(c_i) \mid c_i \in COMPS \setminus \Delta\}$$
$$\cup OBS \cup U_{MEAS} \text{ is consistent}$$

for $\Delta := \{c_i \mid i \in I\}.$
Proof (cont’d).

But, (11) is exactly the condition defining an MBD-diagnosis (see Def. 4). Note, since $N = \emptyset$ by (7), condition (10) is met for any $D$ satisfying (9) and can thus be neglected. Hence, $D = \{\alpha_i \mid i \in I\} \subseteq \mathcal{K}$ is a KBD-diagnosis w.r.t. $kDPI$ iff $\Delta = \{c_i \mid i \in I\} \subseteq \text{COMPS}$ is an MBD-diagnosis for $mDPI$. \qed
Reducing MBD to KBD

Also, there is a bijective correspondence between KBD-conflicts and MBD-conflicts:

**Theorem 2**

Let \( mDPI = (SD, COMPS, OBS, MEAS) \) be an MBD-DPI and \( kDPI = \langle K, B, P, N \rangle_R \) a KBD-DPI modeling \( mDPI \) as per (4) – (8).

Further, let \( COMPS = \{c_1, \ldots, c_n\} \) and \( I \subseteq \{1, \ldots, n\} \). Then, \( C = \{c_i \mid i \in I\} \subseteq COMPS \) is an MBD-conflict for \( mDPI \) iff \( C = \{\alpha_i \mid i \in I\} \subseteq K \) is a KBD-conflict w.r.t. \( kDPI \).

**Proof.**

\( C \) is a KBD-conflict w.r.t. \( kDPI \) iff \( K \setminus C = \{\alpha_i \mid i \in \{1, \ldots, n\} \setminus I\} \) is not a KBD-diagnosis w.r.t. \( kDPI \) (Property 2) iff \( \{c_i \mid i \in \{1, \ldots, n\} \setminus I\} \) is not an MBD-diagnosis for \( mDPI \) (Theorem 1) iff \( \{c_i \mid i \in I\} = C \) is an MBD-conflict for \( mDPI \) ([Reiter, 1987, Prop. 4.2]).
Full-Adder MBD Example (Genesereth, 1984)

A1
X1
X2
A2
O1

Circuit inputs (from top to bottom): 1 0 1
Circuit outputs (from top to bottom): 1 0
Example (cont’d)

\[ \begin{array}{cccc} 
 i & \alpha_i & SD_{beh} & SD_{gen} \\
1 & \neg AB(X_1) \rightarrow beh(X_1) & \bullet & \\
2 & \neg AB(X_2) \rightarrow beh(X_2) & \bullet & \\
3 & \neg AB(A_1) \rightarrow beh(A_1) & \bullet & \\
4 & \neg AB(A_2) \rightarrow beh(A_2) & \bullet & \\
5 & \neg AB(O_1) \rightarrow beh(O_1) & \bullet & \\
6 & out(X_1) = in2(A_2) & \bullet & \\
7 & out(X_1) = in1(X_2) & \bullet & \\
8 & out(A_2) = in1(O_1) & \bullet & \\
9 & in1(A_2) = in2(X_2) & \bullet & \\
10 & in1(X_1) = in1(A_1) & \bullet & \\
11 & in2(X_1) = in2(A_1) & \bullet & \\
12 & out(A_1) = in2(O_1) & \bullet & \\
13 & in1(X_1) = 1 & \bullet & \\
14 & in2(X_1) = 0 & \bullet & \\
15 & in1(A_2) = 1 & \bullet & \\
16 & out(X_2) = 1 & \bullet & \\
17 & out(O_1) = 0 & \bullet & \\
\end{array} \]

COMPS
\[ \{X_1, X_2, A_1, A_2, O_1\} \]

\[ \begin{array}{c} 
 c & beh(c) \text{ for } c \in \text{COMPS} \\
X_1 & out(X_1) = xor(in1(X_1), in2(X_1)) \\
X_2 & out(X_2) = xor(in1(X_2), in2(X_2)) \\
A_1 & out(A_1) = xor(in1(A_1), in2(A_1)) \\
A_2 & out(A_2) = xor(in1(A_2), in2(A_2)) \\
O_1 & out(O_1) = xor(in1(O_1), in2(O_1)) \\
\end{array} \]

<table>
<thead>
<tr>
<th>i</th>
<th>MEAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>\times</td>
</tr>
<tr>
<td>2</td>
<td>\times</td>
</tr>
</tbody>
</table>

MBD-DPI

<table>
<thead>
<tr>
<th>i</th>
<th>\alpha_i</th>
<th>\mathcal{K}</th>
<th>B</th>
</tr>
</thead>
</table>
| 1 | out(X_1) = xor(in1(X_1), in2(X_1)) | \bullet | \\
| 2 | out(X_2) = xor(in1(X_2), in2(X_2)) | \bullet | \\
| 3 | out(A_1) = and(in1(A_1), in2(A_1)) | \bullet | \\
| 4 | out(A_2) = and(in1(A_2), in2(A_2)) | \bullet | \\
| 5 | out(O_1) = or(in1(O_1), in2(O_1)) | \bullet | \\
| 6 | out(X_1) = in2(A_2) | \bullet | \\
| 7 | out(X_1) = in1(X_2) | \bullet | \\
| 8 | out(A_2) = in1(O_1) | \bullet | \\
| 9 | in1(A_2) = in2(X_2) | \bullet | \\
| 10 | in1(X_1) = in1(A_1) | \bullet | \\
| 11 | in2(X_1) = in2(A_1) | \bullet | \\
| 12 | out(A_1) = in2(O_1) | \bullet | \\
| 13 | in1(X_1) = 1 | \bullet | \\
| 14 | in2(X_1) = 0 | \bullet | \\
| 15 | in1(A_2) = 1 | \bullet | \\
| 16 | out(X_2) = 1 | \bullet | \\
| 17 | out(O_1) = 0 | \bullet | \\

KBD-DPI

<table>
<thead>
<tr>
<th>i</th>
<th>p_i \in P</th>
</tr>
</thead>
<tbody>
<tr>
<td>\times</td>
<td>\times</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>n_i \in N</th>
</tr>
</thead>
<tbody>
<tr>
<td>\times</td>
<td>\times</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>r_i \in R</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>consistency</td>
</tr>
</tbody>
</table>

Example (cont’d)
## Example (cont’d)

<table>
<thead>
<tr>
<th>min MBD-conflicts</th>
<th>min KBD-conflicts</th>
</tr>
</thead>
<tbody>
<tr>
<td>{X_1, X_2}, {X_1, A_2, O_1}</td>
<td>{α_1, α_2}, {α_1, α_4, α_5}</td>
</tr>
</tbody>
</table>

### Minimal MBD-conflicts and MBD-diagnoses

<table>
<thead>
<tr>
<th>min MBD-diagnoses</th>
<th>min KBD-diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>{X_1}, {X_2, A_2}, {X_2, O_1}</td>
<td>{α_1}, {α_2, α_4}, {α_2, α_5}</td>
</tr>
</tbody>
</table>

### Minimal KBD-conflicts and KBD-diagnoses
Lessons learned...

- Can find all MBD-diagnoses and MBD-conflicts for any MBD problem by representing it as a KBD problem and solving the latter for KBD-diagnoses and KBD-conflicts, respectively.

- Sequential MBD problem is a special case of Sequential KBD problem: Former can be solved for a given MBD-DPI $MP$ by reducing $MP$ to a KBD-DPI $KP$ and solving latter for $KP$ under the restriction that $N_{new} = \emptyset$.

- Methods targeting KBD-problem are more general than those addressing MBD-problem as they allow the specification of negative information $N_{new}$ in addition to positive one ($P_{new}$ and $MEAS_{new}$, respectively).

- Existing KBD methods such as [Felfernig et al., 2004], [Shchekotykhin et al., 2012] and [Rodler, 2015] are suitable to be used for solving arbitrary MBD problems as per [Reiter, 1987] and [de Kleer & Williams, 1987]
[de Kleer & Williams, 1987] DE KLEER, J., AND WILLIAMS, B. C.
Diagnosing multiple faults.

[Felfernig et al., 2004] FELFERNIG, A., FRIEDRICH, G., JANNACH, D., AND STUMPTNER, M.
Consistency-based diagnosis of configuration knowledge bases.

[Reiter, 1987] REITER, R.
A Theory of Diagnosis from First Principles.

[Rodler, 2015] RODLER, P.
*Interactive Debugging of Knowledge Bases.*

[Shchekotykhin et al., 2012] SHCHEKOTYKHIN, K., FRIEDRICH, G., FLEISS, P., AND RODLER, P.
Interactive Ontology Debugging: Two Query Strategies for Efficient Fault Localization.