

Kalpa Publications in Computing

Volume 4, 2018, Pages 284-296

28th International Workshop on Principles of Diagnosis (DX'17)



Reducing Model-Based Diagnosis to Knowledge Base Debugging

Patrick Rodler and Konstantin Schekotihin

Alpen-Adria Universität, Klagenfurt, Austria firstname.lastname@aau.at

Abstract

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 former can be directly extracted from solutions of the latter. Moreover, we find that 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. As a consequence of these results, KBD approaches can be applied to all systems amenable to MBD.

1 Introduction

Model-Based Diagnosis (MBD) is a general approach to localizing faults in a system of interest, e.g. a digital circuit. The only assumption in MBD is a formal structured description (*model*) of the given system along with observations of the system behavior. The model includes behavioral descriptions of relevant system components (e.g. gates in a digital circuit). The fundamental paradigm underlying MBD is the interaction of prediction and observation. For, in case the observations conflict with the predicted system behavior, MBD aims at the determination of *diagnoses*, explanations of this discrepancy, and possibly the discrimination between multiple diagnoses by gathering additional observations in terms of measurements. The latter process is referred to as *sequential diagnosis* [1].

MBD was originally proposed by Reiter [15] as well as by de Kleer and Williams [1] in 1987. Since then it has generated considerable interest in a great variety of application domains such as physical devices [2, 14, 3], software [22, 12], configuration systems [6], recommender systems [5] and knowledge bases [9, 10, 20, 16]. This has led to a number of different related diagnosis paradigms that build upon MBD and extend, generalize or specialize MBD. In this work we focus on the *Knowledge Base Debugging (KBD)* problem, as studied in e.g. [7, 19, 20, 17, 21, 16]. KBD is basically built on the ideas and concepts from MBD and assumes a (faulty) monotonic knowledge base for which it searches for minimally invasive modifications that yield a solution KB satisfying various postulated properties. The latter are assumed to be given as a set of *requirements* (e.g. logical consistency) and two sets of test cases, the first one (*positive test cases*) including all required entailments of the solution KB and the second one (*negative test cases*) comprising all necessary non-entailments of the solution KB. This test



Figure 1: Full adder MBD example [8, 15] from the domain of circuit diagnosis.

case paradigm can be viewed as an analogon to software debugging, where test cases serve to verify the correctness of the program code.

After recapitulating ideas and concepts from both MBD and KBD, we show that KBD, albeit being a derivative of MBD, is a more general paradigm than MBD. More specifically, we prove that any method solving the KBD problem can be used to solve any MBD problem. In particular, the given proof is constructive, i.e. we specify a mapping from an MBD problem to a KBD problem that constitutes the problem reduction. In addition, we show that there is a direct bijective relationship between the notions of diagnoses and conflicts from both approaches. Finally, we discuss the effects of our analyses on the relationship between *sequential* MBD and KBD problems.

The rest of the work is organized as follows. In Section 2, we revise the general theory of MBD as proposed by [15, 1]. Then, in Section 3 we recap the KBD framework [20, 16]. In Section 4, we demonstrate that (sequential) KBD is a generalization of (sequential) MBD. Finally, Section 5 concludes.

2 Model-Based Diagnosis

We briefly review the classical Model-Based Diagnosis (MBD) problem, as described by [15]. We first characterize a *system*, e.g. a digital circuit or a physical device, which is the subject of a diagnosis task:

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 .

The distinguished unary "abnormal" predicate AB is used in SD to model the expected behavior of components $c \in \text{COMPS}$. Let us denote the first-order sentence describing this expected behavior of c by beh(c) and let $\text{SD}_{beh} := \{\neg AB(c) \rightarrow beh(c) \mid c \in \text{COMPS}\}$. The latter subsumes a statement of the form "if c is nominal (not abnormal), then its behavior is beh(c)" for each system component $c \in \text{COMPS}$. Any behavior different from beh(c) implies that c is at fault, i.e. AB(c) holds. But, an abnormal component does not necessarily manifest a faulty behavior in each situation (*weak fault model* [11, 4]), e.g. for an or-gate c stuck at 1 faulty behavior $\neg beh(c)$ can only be observed if both inputs are 0. Further, SD might include general axioms describing the system domain or descriptions of the interplay between the system components. Let us call the set of these general axioms SD_{gen} . So, $SD = SD_{beh} \cup SD_{gen}$.

The behavior of a system (SD, COMPS) assuming all components working correctly is captured by the description $SD \cup \{\neg AB(c) \mid c \in COMPS\}$. Note, this description is equal to $SD_{gen} \cup \{beh(c) \mid c \in COMPS\}$.

A *diagnosis problem* arises when the observed system behavior – represented by a finite set of first-order sentences OBS – differs from the expected system behavior. Formally, this means that $SD \cup$

 $\{\neg AB(c) \mid c \in COMPS\} \cup OBS \models \bot$. For instance, in circuit diagnosis OBS might be the observation of the system inputs and outputs.

In practical applications there are usually multiple different hypotheses (*diagnoses*) that explain the discrepancy between observed and predicted system behavior. Without additional information about the system, one cannot conjecture a unique diagnosis. Hence, discrimination between diagnoses can be accomplished by means of additional observations MEAS called measurements [15, 1]. Each measurement m in the set of measurements MEAS is a set of first-order sentences [15] describing additional knowledge about the actual system behavior, e.g. whether a particular wire in a faulty circuit is high or low. Usually new measurements are conducted and added to MEAS until some diagnostic goal G is achieved, e.g. the presence of one highly probable or just a single remaining hypothesis. Each added measurement m, if chosen properly, will invalidate some hypotheses.

Formalized, these notions lead to the definitions of an *MBD diagnosis problem instance (MBD-DPI)* and of an *MBD-diagnosis*.¹

Definition 2 (MBD-DPI). Let OBS (system 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 let (SD, COMPS) be a system. Then the tuple (SD, COMPS, OBS, MEAS) is an MBD diagnosis problem instance (MBD-DPI).

Definition 3. Let DPI := (SD, COMPS, OBS, MEAS) be an MBD-DPI and U_{MEAS} denote the union of all $m \in \text{MEAS}$. Then $\text{SD}^*[\Delta] := \text{SD} \cup \{\text{AB}(c) \mid c \in \Delta\} \cup \{\neg \text{AB}(c) \mid c \in \text{COMPS} \setminus \Delta\} \cup \text{OBS} \cup U_{\text{MEAS}}$ for $\Delta \subseteq \text{COMPS}$ 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 $\Delta \subseteq \text{COMPS}$ are faulty and all components in $\text{COMPS} \setminus \Delta$ 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^{*}[Δ] is consistent (i.e. Δ explains OBS and MEAS). An MBD-diagnosis Δ for DPI is called minimal iff there is no MBD-diagnosis Δ' for DPI such that $\Delta' \subset \Delta$.

Sequential MBD addresses the problem of measurement selection in order to reach a predefined diagnostic goal G:

Problem 1 (Sequential MBD).

Given: An MBD-DPI DPI := (SD, COMPS, OBS, MEAS) and a diagnostic goal G.

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

In general, the size of the search space for minimal MBD-diagnoses for (SD, COMPS, OBS, MEAS) is in $O(2^{|\text{COMPS}|})$. A useful concept to restrict this search space is the one of an *MBD-conflict* [15, 1], a set of components whose elements cannot all be healthy given OBS and MEAS:

Definition 5 (MBD-Conflict). Let DPI := (SD, COMPS, OBS, MEAS) be an MBD-DPI. Then $C \subseteq$ COMPS is an MBD-conflict for DPI iff SD $\cup \{\neg AB(c) \mid c \in C\} \cup OBS \cup U_{MEAS}$ is inconsistent. An MBD-conflict C for DPI is called minimal iff there is no MBD-conflict C' for DPI such that $C' \subset C$.

Definition 6 (Hitting Set). Let $S = \{S_1, \ldots, S_n\}$ be a collection of sets and U_S the union of all $S_i \in S$. Then H is called a hitting set of S iff $H \subseteq U_S$ and $H \cap S_i \neq \emptyset$ for all $i = 1, \ldots, n$. A hitting set H of S is minimal iff there is no hitting set H' of S such that $H' \subset H$.

The following result [15] can be used to determine MBD-diagnoses through the computation of MBD-conflicts:

¹The prefix "MBD-" is used to differentiate between MBD concepts and (later discussed) KBD concepts.

Rodler, Schekotihin

Reducing Model-Based Diagnosis to Knowledge Base Debugging

i	$lpha_i$	SD_{beh}	SD_{gen}	OBS	
1	$\neg AB(X_1) \rightarrow beh(X_1)$	•			
2	$\neg AB(X_2) \rightarrow beh(X_2)$	•			
3	$\neg AB(A_1) \rightarrow beh(A_1)$	٠			
4	$\neg AB(A_2) \rightarrow beh(A_2)$	•			
5	$\neg AB(O_1) \rightarrow beh(O_1)$	•			
6	$out(X_1) = in2(A_2)$		•		
7	$out(X_1) = in1(X_2)$		•		
8	$out(A_2) = in1(O_1)$		•		
9	$in1(A_2) = in2(X_2)$		•		
10	$in1(X_1) = in1(A_1)$		•		
11	$in2(X_1) = in2(A_1)$		•		
12	$out(A_1) = in2(O_1)$		•		
13	$in1(X_1) = 1$			•	
14	$in2(X_1) = 0$			•	
15	$in1(A_2) = 1$			•	
16	$out(X_2) = 1$			•	
17	$out(O_1) = 0$			•	
COMPS					
	$\{X_1, X_2, A_1,\}$				
c	$beh(c)$ for $c\in ext{comps}$				
X_1	$out(X_1) = xor(in$	$n1(X_1), n$	$in2(X_1))$		
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	$out(O_1) = xor(in1(O_1), in2(O_1))$				
i	MEAS				
×	>	<			

Table 1: MBD-DPI ExM obtained from the circuit diagnosis problem in Fig. 1.

i	$lpha_i$	\mathcal{K}	B	
1	$out(X_1) = xor(in1(X_1), in2(X_1))$	•		
2	$out(X_2) = xor(in1(X_2), in2(X_2))$ •			
3	$out(A_1) = and(in1(A_1), in2(A_1))$ •			
4	$out(A_2) = and(in1(A_2), in2(A_2))$ •			
5	$out(O_1) = or(in1(O_1), in2(O_1))$ •			
6	$out(X_1) = in2(A_2)$ •			
7	$out(X_1) = in1(X_2)$ •			
8	$out(A_2) = in1(O_1)$ •			
9	$in1(A_2) = in2(X_2)$ •			
10	$in1(X_1) = in1(A_1)$ •			
11	$in2(X_1) = in2(A_1)$ •			
12	$out(A_1) = in2(O_1)$ •			
13	$in1(X_1) = 1$ •			
14	$in2(X_1) = 0$ •			
15	$in1(A_2) = 1$ •			
16	$out(X_2) = 1$ •			
17	$out(O_1) = 0$ •			
i $p_i \in P$				
×	×			
i	$n_i \in N$			
×	×			
i	$r_i \in R$			
1	consistency			
min KBD-conflicts				
$\left\{\alpha_{1},\alpha_{2}\right\},\left\{\alpha_{1},\alpha_{4},\alpha_{5}\right\}$				
min KBD-diagnoses				
	$\left\{ lpha_{1} ight\} ,\left\{ lpha_{2},lpha_{4} ight\} ,\left\{ lpha_{2},lpha_{5} ight\}$			

Table 2: KBD-DPI ExM2K obtained from the MBD-DPI ExM from Tab. 1.

Property 1. A (minimal) MBD-diagnosis for a DPI is a (minimal) hitting set of all minimal MBD-conflicts for this DPI.

Example 1 Let us revisit the circuit diagnosis example given in [15] shown in Fig. 1. The first step towards diagnosing the circuit using MBD is to formulate the problem as an MBD-DPI. The result ExM := (SD, COMPS, OBS, MEAS) is given by Tab. 1 and explained next.

The circuit, i.e. the system to be diagnosed, includes five gates X_1, X_2 (xor-gates), A_1, A_2 (andgates) and O_1 (or-gate), which are at the same time the system components COMPS of interest. The system description $SD = SD_{beh} \cup SD_{gen}$ consists of a knowledge base $SD_{beh} = \{\alpha_1, \ldots, \alpha_5\}$ describing the behavior of each gate given it is working properly, e.g. for gate X_1 , SD_{beh} includes the sentence $\alpha_1 := (\neg AB(X_1) \rightarrow out(X_1) = xor(in1(X_1), in2(X_1)))$. Besides, SD includes a knowledge base $SD_{gen} = \{\alpha_6, \ldots, \alpha_{12}\}$ describing which gate-terminals are connected by wires, e.g. the wire connecting X_1 to X_2 is defined by the sentence $\alpha_7 := (out(X_1) = in1(X_2))$. For simplicity we omit the explicit statement of additional general domain knowledge in SD_{gen} such as axioms for Boolean algebra or axioms restricting wires to only either 0 or 1 values. The observations $OBS = \{\alpha_{13}, \ldots, \alpha_{17}\}$ are given by the system inputs and outputs (see the table in Fig. 1). Finally, since there are no already performed measurements, the set MEAS is empty.

Assuming all components are healthy, i.e. all gates function properly, we find out that $SD^*[\emptyset]$ is inconsistent (cf. Def. 3). That is, the assumption of no faulty components conflicts with the observations OBS made. For instance, if X_1 and X_2 manifest nominal behavior, we can deduce that the output $out(X_2) = 0$ which contradicts the observation sentence $\alpha_{16} := (out(X_2) = 1)$. Supposing either of the components X_1 and X_2 to be nominal, we can no longer deduce $out(X_2) = 0$ (or any other sentence contradicting OBS). Therefore, $C_1 := \{X_1, X_2\}$ is a minimal MBD-conflict (cf. Def. 5). Similarly, we find that $C_2 := \{X_1, A_2, O_1\}$ is the only other minimal MBD-conflict for ExM. Computing minimal hitting sets of all minimal MBD-conflicts C_1, C_2 (Property 1), we obtain three minimal MBD-diagnoses $\Delta_1 := \{X_1\}, \Delta_2 := \{X_2, A_2\}$ and $\Delta_3 := \{X_2, O_1\}$.

Let the diagnostic goal G be the achievement of complete diagnostic certainty, i.e. to single out the correct minimal MBD-diagnosis. The goal of the MBD-problem is then to find new measurements m_1, \ldots, m_k such that there is a single minimal diagnosis Δ for (SD, COMPS, OBS, MEAS $\cup \{m_1, \ldots, m_k\}$). Let the first measurement m_1 be the observation of the terminal $out(X_1)$, and let the value of it be 0. Then, Δ_1 is still a minimal MBD-diagnosis for $\mathsf{ExM}_{new} := (\mathsf{SD}, \mathsf{COMPS}, \mathsf{OBS}, \mathsf{MEAS} \cup \{out(X_1) = 0\}\}$ since the abnormality of X_1 explains both OBS and MEAS. Moreover, all other MBD-diagnoses for ExM_{new} must contain X_1 (since its faultiness is the only explanation for MEAS) and thus be supersets of Δ_1 . Hence, Δ_1 is the only minimal MBD-diagnosis for ExM_{new} and thus the actually faulty component in this scenario is X_1 (under the assumption that a \subseteq -minimal set of components is broken). This fact could be derived by conducting only one measurement.

3 Knowledge Base Debugging

In this section we revisit the Knowledge Base Debugging (KBD) problem [6, 7, 19, 20, 17, 21] (discussed in detail in [16]).

The inputs to a KBD problem can be characterized as follows: Given is a KB \mathcal{K} to be repaired and a KB \mathcal{B} (background knowledge). All sentences in \mathcal{B} are considered correct and all sentences in \mathcal{K} are considered potentially faulty. $\mathcal{K} \cup \mathcal{B}$ does not meet postulated requirements R (where consistency is a least requirement²) or does not feature desired semantic properties, called test cases. Positive test cases (aggregated in the set P) correspond to necessary entailments and negative test cases (aggregated in the set N) represent necessary non-entailments of the correct (repaired) KB (together with the background KB \mathcal{B}). Each test case $p \in P$ and $n \in N$ is a set of sentences. The meaning of a positive test case $p \in P$ is that the union of the repaired KB and \mathcal{B} must entail each sentence (or: the conjunction of sentences) in p, whereas a negative test case $n \in N$ signalizes that some sentence (or: the conjunction of sentences) in n must not be entailed by this union.

The described inputs to the KB debugging problem are captured by the notion of a *KBD diagnosis* problem instance (*KBD-DPI*):

Definition 7 (KBD-DPI). Let

• \mathcal{K} be a KB,

²Consistency is a minimal requirement to a solution KB provided by a debugging system as inconsistency makes a KB completely useless from the semantic point of view.

Rodler, Schekotihin

i	$lpha_i$	\mathcal{K}	\mathcal{B}
1	$\neg H \vee \neg G$	٠	
2	$X\vee F\to H$	٠	
3	$E \to \neg M \wedge X$	٠	
4	$A \to \neg F$	٠	
5	$K \to E$	٠	
6	$C \to B$	٠	
7	$M \to C \wedge Z$	٠	
8	$H \to A$		•
9	$\neg B \vee K$		٠
i	$p_i \in P$		
1	$\{\neg X \to \neg$	$Z\}$	
i	$n_i \in N$		
1	$\{M \to A\}$		
2	$\{E \to \neg G\}$		
3	$\{F \to L\}$		
i	$r_i \in R$		
1	consistency		

Table 3: Example KBD-DPI ExK over propositional logic.

- P, N be sets including sets of sentences,
- $R \supseteq \{consistency\}$ 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).

Example 2 An example ExK of a propositional logic KBD-DPI is depicted by Tab. 3. It includes a KB \mathcal{K} with seven axioms $\alpha_1, \ldots, \alpha_7$, a background KB \mathcal{B} with two axioms α_8, α_9 , one singleton positive test case p_1 and three singleton negative test cases n_1, n_2, n_3 . There is one requirement $r_1 = consistency$ in R imposed on the correct (repaired) KB. It is easy to verify that the standalone KB $\mathcal{B} = \{\alpha_8, \alpha_9\}$ is consistent, i.e. satisfies all $r \in R$, that $\mathcal{K} \cap \mathcal{B} = \emptyset$ and that all involved sets $\mathcal{K}, \mathcal{B}, P$ and N are finite. Moreover, propositional logic is monotonic. Hence, ExK indeed constitutes a KBD-DPI as per Def. 7.

A solution (KB) for a KBD-DPI is characterized as follows:

Definition 8 (Solution KB). Let $DPI := \langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ be a KBD-DPI. Then a KB \mathcal{K}^* is called solution

KB w.r.t. DPI iff all the following conditions hold:

$$\forall r \in R : \mathcal{K}^* \cup \mathcal{B} \text{ fulfills } r \tag{1}$$

$$\forall p \in P : \mathcal{K}^* \cup \mathcal{B} \models p \tag{2}$$

$$\forall n \in N : \mathcal{K}^* \cup \mathcal{B} \not\models n.$$
(3)

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

Remark: The requirements R may include additional (stronger) logical properties apart from consistency. For example, one might postulate solution KBs to be *coherent* [18, 13]. In first-order logic terms (using logic programming notation), a KB \mathcal{K} is coherent iff there is no predicate r in the signature of \mathcal{K} such that $\mathcal{K} \models \forall \mathbf{X} \neg r(\mathbf{X})$ where $\mathbf{X} = X_1, \ldots, X_n$ and n is the arity of the predicate r. That is, coherency means that every predicate in \mathcal{K} can be instantiated without necessarily yielding an inconsistency.

Observing the *Principle of Parsimony* [15], maximal solution KBs \mathcal{K}^* are preferred to non-maximal ones since they result from the input KB \mathcal{K} through the modification of a minimal set of axioms.

Example 3 For the KBD-DPI ExK given by Tab. 3, $\mathcal{K} = \{\alpha_1, \ldots, \alpha_7\}$ is not a solution KB w.r.t. $\langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ since, e.g. clearly $\mathcal{K} \cup \mathcal{B} = \{\alpha_1, \ldots, \alpha_9\} \not\models p_1$ which is a positive test case and therefore has to be entailed. Another reason why $\mathcal{K} = \{\alpha_1, \ldots, \alpha_7\}$ is not a solution KB w.r.t. ExK is that $\mathcal{K} \cup \mathcal{B} \supset \{\alpha_1, \alpha_2, \alpha_3\} \models n_2$, which is a negative test case and hence must not be an entailment. This is straightforward since $\{\alpha_1, \alpha_2, \alpha_3\}$ implies $E \to X$, $X \to H$ and $H \to \neg G$ and thus clearly $n_2 = \{E \to \neg G\}$.

On the other hand, $\mathcal{K}_a^* := \{\} \cup \{Z \to X\}$ is clearly a solution KB w.r.t. ExK as $\{Z \to X\} \cup \mathcal{B}$ is obviously consistent (satisfies all $r \in R$), does entail $p_1 \in P$ and does not entail any $n_i \in N$, $(i \in \{1, 2, 3\})$. However, \mathcal{K}_a^* is not a maximal solution KB since, e.g. $\alpha_5 = (K \to E) \in \mathcal{K}$ can be added to \mathcal{K}_a^* without resulting in the violation of any of (1) - (3).

Maximal solution KBs w.r.t. the given DPI are, e.g. $\mathcal{K}_b^* := \{\alpha_1, \alpha_4, \alpha_5, \alpha_6, \alpha_7, p_1\}$ (resulting from the deletion of $\{\alpha_2, \alpha_3\}$ from \mathcal{K} and the addition of p_1) or $\mathcal{K}_c^* := \{\alpha_1, \alpha_2, \alpha_5, \alpha_6, p_1\}$ (resulting from the deletion of $\{\alpha_1, \alpha_4, \alpha_7\}$ from \mathcal{K} and the addition of p_1). That these KBs constitute solution KBs can be verified by checking the three conditions named by Def. 8. Indeed, adding an additional axiom in \mathcal{K} to any of the two KBs leads to the entailment of a negative test case $n \in N$. That is, no solution KBs can contain a proper superset of the axioms from \mathcal{K} that are contained in any of the two solution KBs \mathcal{K}_b^* and \mathcal{K}_c^* . Hence, both are maximal.

In general, there is not a unique (maximal) solution KB resulting from the deletion of one and the same set of axioms \mathcal{D} from the original KB \mathcal{K} :

Example 4 Consider again ExK in Tab. 3 and assume $\mathcal{D} = \{\alpha_2, \alpha_3\}$ is deleted from \mathcal{K} . Then one solution KB constructible from $\mathcal{K} \setminus \mathcal{D}$ is \mathcal{K}_b^* given in the last example. To determine the maximal solution KB \mathcal{K}_b^* from $\mathcal{K} \setminus \mathcal{D}$, the most straightforward way of adding just all sentences occurring in positive test cases in P has been chosen in this case. Other maximal solution KBs obtainable from adding sentences to $\mathcal{K} \setminus \mathcal{D}$ are, e.g. $\mathcal{K}_{b1}^* := \{\alpha_1, \alpha_4, \alpha_5, \alpha_6, \alpha_7, Z \to X\}$ (which differs syntactically, but not semantically from \mathcal{K}_b^*) and $\mathcal{K}_{b2}^* := \{\alpha_1, \alpha_4, \alpha_5, \alpha_6, \alpha_7, Z \to X \land W\}$ (which differs both syntactically and semantically from \mathcal{K}_b^* yielding the entailment $Z \to W$ which is not implied by \mathcal{K}_b^*).

Since (i) the diagnostic evidence of a DPI in terms of positive test cases P does not justify the inclusion of sentences (semantically) different from U_P (cf. [7, 20, 16]) and (ii) in KBD applications

only one solution KB is sought, $\mathcal{K} \setminus \mathcal{D} \cup U_P$ is regarded as 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.

A *KBD-diagnosis* is defined in terms of the axioms \mathcal{D} that must be deleted from the KB \mathcal{K} of a DPI in order to construct a solution KB w.r.t. this DPI. In particular, the deletion of \mathcal{D} from \mathcal{K} targets the fulfillment of (1) and (3) such that U_P can be added to the resulting modified KB $\mathcal{K} \setminus \mathcal{D}$ without introducing any new violations of (1) or (3).

Definition 9 (KBD-Diagnosis). Let DPI := $\langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ be a KBD-DPI. A set of sentences $\mathcal{D} \subseteq \mathcal{K}$ is called a KBD-diagnosis w.r.t. DPI iff $(\mathcal{K} \setminus \mathcal{D}) \cup U_P$ is a solution KB w.r.t. DPI (i.e. $\mathcal{K}^* := (\mathcal{K} \setminus \mathcal{D}) \cup U_P$ satisfies (1) – (3)). A KBD-diagnosis \mathcal{D} w.r.t. DPI is minimal iff there is no $\mathcal{D}' \subset \mathcal{D}$ such that \mathcal{D}' is a KBD-diagnosis w.r.t. DPI.

Remark: Since $(\mathcal{K} \setminus \mathcal{D}) \cup U_P$ trivially satisfies (2) due to the inclusion of U_P , \mathcal{D} is a KBD-diagnosis w.r.t. DPI iff $\mathcal{K}^* := (\mathcal{K} \setminus \mathcal{D}) \cup U_P$ satisfies (1) and (3).

The relationship between maximal canonical solution KBs and minimal KBD-diagnoses w.r.t. a DPI is as follows (cf. [16, p. 34]):

Property 2. Let DPI be a KBD-DPI. Then the set of all maximal canonical solution KBs w.r.t. DPI is given by $\{(\mathcal{K} \setminus \mathcal{D}) \cup U_P \mid \mathcal{D} \text{ is a minimal KBD-diagnosis w.r.t. DPI}\}.$

Therefore, KBD methods focus on the computation of minimal KBD-diagnoses in order to find all maximal canonical solution KBs.

In a completely analogous way as MBD-conflicts provide an effective mechanism for focusing the search for MBD-diagnoses, KBD-conflicts are exploited for KBD-diagnoses calculation. Simply put, a (minimal) KBD-conflict is a (minimal) per se faulty subset of the original KB \mathcal{K} , i.e. one source causing the faultiness of \mathcal{K} in the context of $\mathcal{B} \cup U_P$. For a KBD-conflict there is no extension that yields a solution KB. Instead, such an extension is only possible after deleting appropriate axioms from the KBD-conflict.

Definition 10 (KBD-Conflict). Let DPI := $\langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ be a KBD-DPI. A set of formulas $\mathcal{C} \subseteq \mathcal{K}$ is called a KBD-conflict w.r.t. DPI iff $\mathcal{C} \cup U_P$ is not a solution KB w.r.t. DPI (i.e. $\mathcal{K}^* := \mathcal{C} \cup U_P$ violates at least one of (1) - (3)). A KBD-conflict \mathcal{C} w.r.t. DPI is minimal iff there is no $\mathcal{C}' \subset \mathcal{C}$ such that \mathcal{C}' is a KBD-conflict w.r.t. DPI.

Property 3. [7, Prop. 2] Let DPI be a KBD-DPI. Then a (minimal) KBD-diagnosis w.r.t. DPI is a (minimal) hitting set of all minimal conflict sets w.r.t. DPI.

Property 4. [16, Prop. 3.4] Let $DPI := \langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ be a KBD-DPI. Then a KBD-diagnosis w.r.t. DPI exists iff $\mathcal{B} \cup U_P$ satisfies all $r \in R$ and $\mathcal{B} \cup U_P \not\models n$ for all $n \in N$.

Example 5 A list of all minimal KBD-conflicts w.r.t. the example DPI ExK (see Tab. 3) is enumerated in Tab. 4. In Example 3, we gave an explanation why C_1 is a KBD-conflict (violation of $n_2 \in N$, cf. the explanation in column 3 of Tab. 4). C_1 is minimal since, first, it is consistent, i.e. satisfies all $r \in R$, and does not entail any of the negative test cases n_1, n_3 . Consequently, by logical monotonicity no proper subset of C_1 can violate r, n_1 or n_3 . Second, the elimination of any axiom $\alpha_i (i \in \{1, 2, 3\})$ from C_1 breaks the entailment of the negative test case n_2 .

 $C_2 := \{\alpha_2, \alpha_4\}$ is a KBD-conflict due to the fact that $\alpha_2 \equiv \{X \to H, F \to H\}$ together with $\alpha_8 (\in \mathcal{B}) = H \to A$ and $\alpha_4 = A \to \neg F$ clearly yields $F \to \neg F \equiv \neg F$ which, in particular, implies $n_3 = \{F \to L\} \equiv \{\neg F \lor L\}$.

Rodler, Schekotihin

min KBD-conflict X	$\{i \alpha_i \in X\}$	explanation
\mathcal{C}_1	$\{1, 2, 3\}$	$\models n_2$
\mathcal{C}_2	$\{2, 4\}$	$\cup \{8\} \models \neg F \ (\models n_3)$
\mathcal{C}_3	$\{2,7\}$	$\cup \{p_1, 8\} \models n_1$
\mathcal{C}_4	$\{3, 5, 6, 7\}$	$\cup \{9\} \models \neg M \ (\models n_1)$
min KBD-diagnosis X	$\{i \alpha_i \in X\}$	explanation
\mathcal{D}_1	$\{2, 3\}$	Property 1
\mathcal{D}_2	$\{2, 5\}$	Property 1
\mathcal{D}_3	$\{2, 6\}$	Property 1
${\cal D}_4$	$\{2,7\}$	Property 1
\mathcal{D}_5	$\{1, 4, 7\}$	Property 1
\mathcal{D}_6	$\{3, 4, 7\}$	Property 1

Table 4: Minimal KBD-conflicts and KBD-diagnoses for the KBD-DPI ExK in Tab. 3.

The verification of the minimality of C_2 and the fact that C_3 and D_4 are (all) other minimal KBDconflicts works analogous and is straightforward under consideration of the explanations in column 3 of Tab. 4.

All minimal KBD-diagnoses $\mathcal{D}_1, \ldots, \mathcal{D}_6$ w.r.t. ExK (see Tab. 4) result from a minimal hitting set computation over all minimal KBD-conflicts.

The relationship between the notions *KBD-diagnosis*, solution *KB* and *KBD-conflict* is as follows:

Property 5. [16, Cor. 3.3] Let $\mathcal{D} \subseteq \mathcal{K}$. Then the following statements are equivalent:

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

Example 6 Since, e.g., $\mathcal{K} \setminus \mathcal{D} := \{\alpha_1, \alpha_2\}$ is not a KBD-conflict w.r.t. ExK (Tab. 3), we obtain that $\mathcal{D} = \mathcal{K} \setminus (\mathcal{K} \setminus \mathcal{D}) = \{\alpha_1, \dots, \alpha_7\} \setminus \{\alpha_1, \alpha_2\} = \{\alpha_3, \dots, \alpha_7\}$ is a KBD-diagnosis w.r.t. ExK, albeit not a minimal one (α_5 as well as α_6 can be deleted from it while preserving its KBD-diagnosis property). Further on, $(\mathcal{K} \setminus \mathcal{D}) \cup U_P = \{\alpha_1, \alpha_2, p_1\}$ must be a solution KB w.r.t. ExK.

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.

4 Reducing Model-Based Diagnosis to Knowledge Base Debugging

We next demonstrate that the classical MBD problem described in Sec. 2 can be reduced to the KBD problem explicated in Sec. 3. That is, any MBD-DPI can be modeled as a KBD-DPI, and the solutions

Rodler, Schekotihin

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. Moreover, all MBD-diagnoses for mDPI can be computed from the KBD-diagnoses w.r.t. kDPI.

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

$$\mathcal{K} = \{ \alpha_i \mid \alpha_i := beh(c_i), c_i \in \text{COMPS} \}$$
(4)

$$\mathcal{B} = \text{OBS} \cup \text{SD}_{gen} \tag{5}$$

$$P = \text{MEAS} \tag{6}$$

$$N = \emptyset \tag{7}$$
$$R = \{ consistency \} \tag{8}$$

$$R = \{consistency\}\tag{8}$$

That is, \mathcal{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. 9 and the subsequent remark, $\mathcal{D} \subseteq \mathcal{K}$ is a KBD-diagnosis for kDPI iff both

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

and

$$(\mathcal{K} \setminus \mathcal{D}) \cup \mathcal{B} \cup U_P \not\models n \text{ for all } n \in N$$
(10)

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

Using (4) – (8) above, condition (9) for \mathcal{D} is equivalent to the consistency of $SD_{beh} \cup \{AB(c_i) \mid i \in I\} \cup \{\neg AB(c_i) \mid i \in \{1, ..., n\} \setminus I\} \cup OBS \cup SD_{gen} \cup U_{MEAS}$ which in turn yields that

$$SD \cup \{AB(c_i) \mid c_i \in \Delta\} \cup \{\neg AB(c_i) \mid c_i \in COMPS \setminus \Delta\} \cup OBS \cup U_{MEAS} \text{ is consistent}$$
(11)

for $\Delta := \{c_i \mid i \in I\}$. 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 \mathcal{D} satisfying (9) and can thus be neglected. Hence, $\mathcal{D} = \{\alpha_i \mid i \in I\} \subseteq \mathcal{K}$ is a KBD-diagnosis w.r.t. kDPI iff $\Delta = \{c_i \mid i \in I\} \subseteq$ COMPS is an MBD-diagnosis for mDPI.

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 $kDPI = \langle \mathcal{K}, \mathcal{B}, P, N \rangle_R$ 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 $C = \{\alpha_i \mid i \in I\} \subseteq \mathcal{K}$ is a KBD-conflict w.r.t. kDPI.

Proof. C is a KBD-conflict w.r.t. kDPI iff $\mathcal{K} \setminus C = \{\alpha_i \mid i \in \{1, ..., n\} \setminus I\}$ is not a KBD-diagnosis w.r.t. kDPI (Property 5) iff $\{c_i \mid i \in \{1, ..., n\} \setminus I\}$ is not an MBD-diagnosis for mDPI (Theorem 1) iff $\{c_i \mid i \in I\} = C$ is an MBD-conflict for mDPI ([15, Prop. 4.2]).

Let us exemplify these theoretical results:

Example 7 Reconsider the circuit diagnosis example of Fig. 1. The formalization of the circuit problem as an MBD-DPI ExM was already discussed in Example 1. The formulation of this MBD-DPI as a KBD-DPI ExM2K as per (4) – (8) is depicted by Tab. 2. All minimal KBD-conflicts and their minimal hitting sets, i.e. the minimal KBD-diagnoses (Property 3), are given in the lower part of Tab. 2. For instance, $C = \{\alpha_1, \alpha_4, \alpha_5\}$ is a KBD-conflict w.r.t. ExM2K since $C \cup B \cup U_P \models \bot$. We briefly sketch why this holds. $\alpha_{13} \in B = (in1(X_1) = 1)$, $\alpha_{14} \in B = (in2(X_1) = 0)$ and $\alpha_1 = (out(X_1) = xor(in1(X_1), in2(X_1)))$ imply that $out(X_1) = xor(1, 0) = 1$, which, along with $\alpha_6 \in B = (out(X_1) = in2(A_2))$, entails $in2(A_2) = 1$, which in turn, together with $\alpha_{15} \in B = (in1(A_2) = 1)$ and $\alpha_4 = (out(A_2) = and(in1(A_2), in2(A_2)))$, lets us deduce that $out(A_2) = and(1, 1) = 1$. Because of $\alpha_8 \in B = (out(A_2) = in1(O_1))$ we have that $in1(O_1) = 1$ which yields $out(O_1) = or(1, in2(O_1)) = 1$ due to $\alpha_5 = (out(O_1) = or(in1(O_1), in2(O_1)))$. However, $\alpha_{17} \in B$ states that $out(O_1) = 0$, a contradiction.

C is minimal since all elements of C – namely, the behavioral descriptions beh(g) of gates $g \in \{X_1, A_2, O_1\}$ (see Tab. 2) – were necessary to derive the outlined contradiction. In fact, no proper subset of C can be used to deduce any negative test case (trivially, as the set N is empty) or any contradiction (possibly different from the one given above). Intuitively, the latter holds since any $C' \subset C$ includes too few behavioral descriptions of components so that there is no "open" path for constraint propagation from inputs to outputs of the circuit. C itself, on the other hand, enables to propagate information from all three inputs via gates X_1 , A_2 and O_1 towards the second output. By the fact that C is a minimal KBD-conflict, Theorem 2 lets us conclude that $\{X_1, A_2, O_1\}$ is a minimal MBD-conflict for the original MBD-problem stated in Tab. 1.

What becomes nicely evident at this point is the principle of transformation between MBD and KBD. Whereas in MBD behavioral descriptions of components are "disabled" via abnormality assumptions about components, in KBD it is exactly these descriptions that make up the KB, and they are "inactivated" by just deleting them from the KB.

The justification for the minimal KBD-conflict $\{\alpha_1, \alpha_2\}$ follows essentially the same argumentation as was given in Example 1 to explain C_1 . In this case, Theorem 2 yields that $\{X_1, X_2\}$ must be a minimal MBD-conflict for the original MBD-problem because $\alpha_1 = beh(X_1)$ and $\alpha_2 = beh(X_2)$.

Computing minimal hitting sets of the minimal KBD-conflicts $\{\alpha_1, \alpha_4, \alpha_5\}$ and $\{\alpha_1, \alpha_2\}$ yields all minimal KBD-diagnoses w.r.t. ExM2K given at the bottom of Tab. 2. Application of Theorem 1 lets us conclude that $\{X_1\}$ (due to the minimal KBD-diagnosis $\{\alpha_1\}$), $\{A_2, X_2\}$ (due to $\{\alpha_4, \alpha_2\}$) and $\{O_1, X_2\}$ (due to $\{\alpha_5, \alpha_2\}$) are exactly the minimal MBD-diagnoses for the orginal MBD-problem. \Box

5 Conclusions

To sum up, we can find all MBD-diagnoses and MBD-conflicts for any MBD problem by representing it as a KBD-DPI and solving the latter for KBD-diagnoses (Theorem 1) and KBD-conflicts (Theorem 2), respectively. Considering the sequential problems associated with MBD (Problem 1) and KBD (Problem 2), the derived results indicate that the former is a special case of the latter. In particular, Problem 1 can be solved for a given MBD-DPI MP by reducing MP to a KBD-DPI KP as per Theorem 1 and solving Problem 2 for KP under the restriction that $N_{new} = \emptyset$. In other words, sequential KBD methods specifying only positive test cases allow to solve the sequential MBD problem. This shows that methods targeting Problem 2 are more general than those addressing Problem 1 as they allow the specification of negative information N_{new} in addition to positive one (P_{new} and MEAS_{new}, respectively). From the practical point of view, our findings imply that existing KBD methods such as [6, 7, 19, 20, 17, 21, 16] are in principle eligible to be used for solving arbitrary MBD problems as per [15, 1].

Acknowledgments

This work was supported by the Carinthian Science Fund (KWF), contract KWF-3520/26767/38701.

References

- Johan de Kleer and Brian C. Williams. Diagnosing multiple faults. Artificial Intelligence, 32(1):97–130, 1987.
- [2] Oskar Dressler and Peter Struss. The consistency-based approach to automated diagnosis of devices. *Principles of Knowledge Representation*, pages 269–314, 1996.
- [3] Alexander Feldman, Gregory M. Provan, and Arjan J. C. van Gemund. A model-based active testing approach to sequential diagnosis. J. Artif. Intell. Res. (JAIR), 39:301–334, 2010.
- [4] Alexander Feldman, Gregory M. Provan, and Arjan J.C. van Gemund. Solving strong-fault diagnostic models by model relaxation. In *IJCAI'09*, pages 785–790, 2009.
- [5] Alexander Felfernig, Gerhard Friedrich, Karl Isak, Kostyantyn Shchekotykhin, Erich Teppan, and Dietmar Jannach. Automated debugging of recommender user interface descriptions. *Applied Intelligence*, 31(1):1– 14, 2009.
- [6] Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, and Markus Stumptner. Consistency-based diagnosis of configuration knowledge bases. Artificial Intelligence, 152(2):213 – 234, 2004.
- [7] Gerhard Friedrich and Kostyantyn Shchekotykhin. A General Diagnosis Method for Ontologies. In ISWC'05, pages 232–246, 2005.
- [8] Michael R. Genesereth. The use of design descriptions in automated diagnosis. Artificial Intelligence, 24(1-3):411–436, 1984.
- [9] Aditya Kalyanpur. *Debugging and Repair of OWL Ontologies*. PhD thesis, University of Maryland, College Park, 2006.
- [10] Aditya Kalyanpur, Bijan Parsia, Matthew Horridge, and Evren Sirin. Finding all Justifications of OWL DL Entailments. In *ISWC'07*, pages 267–280, 2007.
- [11] Johan de Kleer, Alan K. Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56, 1992.
- [12] Cristinel Mateis, Markus Stumptner, Dominik Wieland, and Franz Wotawa. Model-Based Debugging of Java Programs. In AADEBUG'00, 2000.
- [13] Bijan Parsia, Evren Sirin, and Aditya Kalyanpur. Debugging OWL ontologies. In WWW'05, pages 633–640, 2005.
- [14] Jurryt Pietersma, Arjan J.C. van Gemund, and Andre Bos. A model-based approach to sequential fault diagnosis. In *IEEE Autotestcon*, 2005, pages 621–627, 2005.
- [15] Raymond Reiter. A Theory of Diagnosis from First Principles. Artificial Intelligence, 32(1):57–95, 1987.
- [16] Patrick Rodler. Interactive Debugging of Knowledge Bases. PhD thesis, Alpen-Adria Universität Klagenfurt, 2015. http://arxiv.org/pdf/1605.05950v1.pdf.
- [17] Patrick Rodler, Kostyantyn Shchekotykhin, Philipp Fleiss, and Gerhard Friedrich. RIO: Minimizing User Interaction in Ontology Debugging. In Web Reasoning and Rule Systems, pages 153–167. 2013.
- [18] Stefan Schlobach, Zhisheng Huang, Ronald Cornet, and Frank Harmelen. Debugging Incoherent Terminologies. *Journal of Automated Reasoning*, 39(3):317–349, 2007.
- [19] Kostyantyn Shchekotykhin and Gerhard Friedrich. Query strategy for sequential ontology debugging. In ISWC'10, pages 696–712, 2010.
- [20] Kostyantyn Shchekotykhin, Gerhard Friedrich, Philipp Fleiss, and Patrick Rodler. Interactive Ontology Debugging: Two Query Strategies for Efficient Fault Localization. Web Semantics: Science, Services and Agents on the World Wide Web, 12-13:88–103, 2012.

- [21] Kostyantyn Shchekotykhin, Gerhard Friedrich, Patrick Rodler, and Philipp Fleiss. Sequential diagnosis of high cardinality faults in knowledge-bases by direct diagnosis generation. In *ECAI'14*, pages 813–818, 2014.
- [22] Markus Stumptner and Franz Wotawa. Debugging functional programs. In IJCAI'99, pages 1074–1079, 1999.