

Model-Based Debugging using Multiple Abstract Models

Wolfgang Mayer*, Markus Stumptner*¹

* *University of South Australia
Advanced Computing Research Centre
Mawson Lakes, Adelaide SA 5095, Australia*

ABSTRACT

This paper introduces an automatic debugging framework that relies on model-based reasoning techniques to locate faults in programs. In particular, model-based diagnosis, together with an abstract interpretation based conflict detection mechanism is used to derive diagnoses, which correspond to possible faults in programs. Design information and partial specifications are applied to guide a model revision process, which allows for automatic detection and correction of structural faults.

KEYWORDS: Model-based Debugging, Diagnosis, Abstract Interpretation, Program Analysis

1 Introduction

Detecting a faulty behavior within a program, locating the cause of the fault, and fixing the fault by means of changing the program, continues to be a crucial and challenging task in software development. Many papers have been published so far in the domain of detecting faults in software, e.g., testing or formal verification [CDH⁺00], and locating them, e.g., program slicing [Wei84] and automatic program debugging [Llo87]. More recently model-based diagnosis [Rei87] has been used for locating faults in software [CFD93, MSWW02a].

This paper extends previous research in several directions: Firstly, a parameterized debugging framework is introduced, which integrates dynamic and static properties, as well as design information of programs. The framework is based on results derived in the field of abstract interpretation [CC77], and can therefore be parameterized with different lattices and context selection strategies.

Secondly, the one-to-one correspondence between model components and program statements is replaced by a hierarchy of components, which provides means for more efficient reasoning procedures, as well as more flexibility when focusing on interesting parts of a program.

This work is organized as follows. In Section 2, we give an introduction to model-based debugging. Section 3 describes mapping from source code to model components and the (approximate) computation of program effects in the style of [CC77] and [Bou93]. The next section discusses the modeling of programs and the reasoning framework. In Section 5, we provide an example which puts together the different models and demonstrates the debugging capabilities of our approach.

In M. Ronsse, K. De Bosschere (eds), proceedings of the Fifth International Workshop on Automated Debugging (AADE-BUG 2003), September 2003, Ghent. Computer Research Repository (<http://www.acm.org/corr/>), cs.SE/yymmnnn; whole proceedings: cs.SE/0309027.

¹E-mail: {mayer,mst}@cs.unisa.edu.au

Section 6 provides details about our implementation. Finally, we discuss related work and conclude the paper.

2 Model-based Debugging

To locate faults using model-based reasoning techniques, the source code of the program P to be analyzed must be available. Also, a set of test cases \mathcal{TC} is required, which (partially) specify the expected behavior of P . Test cases can be as simple as a set of input-output vectors or even just a list of correct and incorrect output variables. The connection to the model-based diagnostic framework is realized through a set \mathcal{COMPS} and a set \mathcal{M} of models. \mathcal{COMPS} contains the set of components of which fault candidates are composed, whereas each $m \in \mathcal{M}$ describes the program behavior (possibly at an abstract level) and is used to detect discrepancies between the expected and the obtained behavior of the program. A fault candidate is a part of P 's source code that, when assumed to show arbitrary effects, does not conflict with any test case in \mathcal{TC} any more. A fault candidate conflicts with a test case t if the modified program corresponding to the fault candidate derives values different from the ones specified in t . A model $m \in \mathcal{M}$ of P is a (partial) description of the P 's behavior, derived automatically from the source code of P .

For example, using a model that describes dependencies between components, where each component corresponds to a statement, the (faulty) program

```

1  int r = 3;
2  float area = r*3.141f;
3  float circ = 2.f*r*3.141f;

```

can be described as follows.

If statement 1 is correct, the value of `r` is correct. If statement 2 and `r` are correct, `area` is correct, too. `circ` is correct provided statement 3 and `r` are correct.

Translated to first order logic, this can be represented as follows:

$$\begin{aligned}
& (\neg ab(c_1) \rightarrow correct(r)) \wedge \\
& (\neg ab(c_2) \wedge correct(r) \rightarrow correct(area)) \wedge \\
& (\neg ab(c_3) \wedge correct(r) \rightarrow correct(circ)).
\end{aligned}$$

c_1 to c_3 represent the components corresponding to the statements in lines 1 to 3, respectively, and *correct* is a predicate that asserts that the variable passed as argument has the correct value (specific to the test case under consideration). Test cases are represented as conjunctions of *correct* literals. For example, $correct(circ) \wedge \neg correct(area)$ expresses that after running the program, variable `circ` is correct, whereas `area` is incorrect. *ab* is used by the diagnostic engine to disable the model of certain components and check if the remaining model is still inconsistent with the test case. A more formal elaboration can be found below.

We recall some of the basic definitions from model-based diagnosis [Rei87], slightly adapted for our purposes:

Definition 1 (Diagnosis Problem) *A diagnosis problem is a triple $(SD, \mathcal{COMPS}, \mathcal{OBS})$ where SD is the system description, \mathcal{COMPS} is the set of components in SD , and \mathcal{OBS} is the set of observations.*

Here, $SD \in \mathcal{M}$ is a model of P , and $\mathcal{OBS} \in \mathcal{TC}$ is the information specified by test cases. Note that \mathcal{OBS} contains the *expected* result of test cases, not the actual result obtained from the faulty program. Also, \mathcal{OBS} is not restricted to pure input and output specifications; intermediate results can also be checked using assertions (see Section 3.1).

Definition 2 (Diagnosis) A set $\Delta \subseteq COMPS$ is a diagnosis for a diagnosis problem $(SD, COMPS, OBS)$ iff $SD \cup OBS \cup \{\neg ab(C) \mid C \in COMPS \setminus \Delta\}$ is consistent, where $ab(C)$ denotes that component C is not working as specified in SD .

Each component $C \in COMPS$ corresponds to a part of P and therefore, components in Δ indicate possible faults in the program. The $\neg ab(C)$ behavior of a component C is an abstraction [CC77] of the semantics of the code fragment represented by C , as given by the language specification. The $ab(C)$ behavior denotes a possible fault and generally permits arbitrary effects.

Diagnoses can be computed efficiently using the concept of conflicts, which are sets of components that cannot be all functioning correctly without contradicting at least one $t \in TC$.

Definition 3 (Conflict) $\Delta \subseteq COMPS$ is a conflict for $(SD, COMPS, OBS)$ iff $SD \cup OBS \cup \{\neg ab(C) \mid C \in \Delta\}$ is inconsistent.

The basic principle of MBD is to use \mathcal{M} to derive conflicts given TC as observations. The conflicts are then used by the diagnostic engine to compute diagnoses, which are mapped back to the program's source code to indicate possible faults. To minimize the number of fault candidates, we are only interested in subset-minimal diagnoses, which can be derived from subset-minimal conflicts [Rei87].

Revisiting the previous example, it is easy to see that $\{ab(c_3)\}$ cannot be a diagnosis, as the model derives the conflict $\{\neg ab(c_1), \neg ab(c_2)\}$. However, $\{ab(c_1)\}$ and $\{ab(c_2)\}$ are both diagnoses. $\{ab(c_1) \wedge ab(c_3)\}$ is also a diagnosis, but not subset-minimal, as it contains the diagnosis $\{ab(c_1)\}$.

As a possible extension not covered in this paper, the approach could be extended to output the most likely diagnoses, given prior probabilities for each component. These probabilities can be obtained by counting the number of correct and faulty test cases that the statements corresponding to each component are executed in [Tip95, MSWW02b, JHS02].

3 Modeling Program Behavior

A key aspect of every MBD system is the construction of the set \mathcal{M} of models and the mapping between the program and \mathcal{M} s components.

Previous work [MSW00, MSWW02a] derives the models from the source code without considering runtime information, which often results in large and complex models. We construct the models dynamically, which, by exploiting runtime information, can lead to smaller and more concise models.

Another limitation imposed by these earlier modeling approaches is the representation of every statement and (sub-)expression in P as a separate component in the model. Even though these models allow for very detailed reasoning, this is rarely required in practice and leads to a large number of diagnoses and to increased computational requirements.

To overcome these limitations, we employ an iterative, hierarchical diagnostic process, where the mapping from P to $COMPS$ is refined incrementally (starting with a single component for each method), depending on the results of previous diagnostic analysis (see [FFJS00] for a similar approach).

Previous models behave poorly when the number of loop iterations or recursion depth is not known in advance. The combination of static program analysis and dynamic execution proposed in the next sections provides an effective combination, which is well-suited for dealing with such constructs.

3.1 Approximate Program Analysis

Static program analysis, in particular Abstract Interpretation [CC77], has successfully been applied to derive properties of programs, even in the absence of specific test cases. Also, the framework is customizable with different abstractions of the concrete semantics of a program.

We recall the basic definitions of Abstract Interpretation, as given in [CC77, Bou93]:

The mapping from the concrete semantics, represented as a lattice $(\mathcal{P}(S), \emptyset, S, \subseteq, \cup, \cap)$ (S denotes the set of program states), to the abstract, finitely represented lattice, $(\mathcal{P}^\#(S), \perp, \top, \sqsubseteq, \sqcup, \sqcap)$, is given by a Galois Connection (α, γ) , where α maps sets of states to their best approximation, and γ maps every abstract property to its meaning in $\mathcal{P}(S)$.

The approximate semantics of a program P can then be expressed as fixpoint over a set \mathcal{X} of equations derived from P 's source code. The equations are composed of abstract operations $\Phi^\# \equiv \gamma \circ \Phi \circ \alpha$, which model the effects of every operation Φ in P . An approximation of the forward semantics is given by the solution of $\text{lfp } \lambda X \cdot (E \sqcap \mathcal{X}(X))$ (starting at \perp), where E denotes the approximation of the entry states. In case the abstract lattice is of infinite height, narrowing and widening operators have to be applied to ensure termination of the computation. For a more in-depth discussion see [CC77]. Bourdoncle described similar approximations of backward semantics and added intermittent and invariant assertions for program analysis [Bou93].

To incorporate intermittent (“sometimes”) and invariant (“always”) assertions into the analysis, a sequence of forward and backward reasoning steps can be defined to approximate the entry and exit states which guarantee the validity of the assertions [Bou93]. Intermittent assertions express conditions that must eventually hold in each program execution, but not necessarily each time the program point is reached. Invariant assertions on the other hand have to be true every time the corresponding program point is reached (if it is reached at all). For example, the assertion `sometime true`; at the end of a program asserts that the program must eventually terminate. Similarly, `always i >= 0 && i <= 10` asserts that whenever that point of the program is reached, i must be between 0 and 10. Note that *sometime* and *always* assertions, contrary to what the names may imply, do not require the presence of multiple test cases. For example, consider a test case where a loop executes multiple iterations. In this case, the difference between `sometime C` and `always C` is evident: `always C` requires condition C to be true in *every* iteration, whereas `sometime C` only requires that for *one* iteration.

3.2 Avoiding Imprecision

The approximation of complex programs leads to possible imprecision, which is undesirable for automatic debugging. In particular, (1) aliasing between variables has to be approximated, (2) it can be difficult to derive useful properties for arrays, and (3) partitioning of the domain of the abstraction function severely impacts the outcome of the analysis. Further imprecision may be introduced by composition of the abstractions for each statement. To deal with (1) and (2), numerous different abstractions [HP00] and partial evaluation approaches [Col97] have been developed. However, they are generally not very well-suited for MBD, because the results are often too imprecise to derive a conflict. To overcome (3), [Bou92] introduced a model that is able to refine the domain based on the current partitioning. However, even in this framework, the choice of approximation operators remains crucial (and program dependent).

To circumvent the aforementioned shortcomings, we employ the information from test cases to avoid approximation whenever possible, and rely on static analysis only as a fallback in case the program's behavior is only partially specified or exceeds user-defined bounds. A more detailed discussion is provided in Section 4.

4 Model Construction

In this section, we present a model that follows the execution semantics of the program. Based on the semantic approximation introduced in the previous section, a separate model for each test case is constructed by abstract interpretation of the program, using the entry state specified by the test case. Test case information (pre- and postconditions values, as well as intermediate assertions) is mapped to *sometime* and *always* assertions. This differs from traditional abstract interpretation techniques [CC77] as we generate the equations representing the system dynamically while the fixpoint

is computed, which is advantageous when combined with the MBD engine and partitioning strategies (see below). The model derives a contradiction iff there exists no feasible path between the entry state and the exit state of the program.² To determine the set of components the conflict is composed of, we follow the approach of [MT02]. The algorithm can be summarized as follows. After a conflict has been detected, the derivation tree is analyzed to find the subset-minimal set of constraint needed to derive the inconsistency. This is done by recursively subdividing the derivation tree and pruning sets that cannot contribute to the conflict.

The dynamic approach, together with the test case information allows us to explore only these parts of the model which may actually be executed. Especially for object-oriented languages like Java, with many possibly exception-throwing statements, this approach results in significantly smaller models. For example, if a branch of a conditional can be eliminated, its statement need not be considered and the data flow ϕ and σ functions [Ana99] can be eliminated, too.

4.1 Partitioning Strategies

Crucial to the accuracy of the results is the selection of partitioning strategies for contexts of method calls. This corresponds to the selection of widening operators in [Bou92]. We propose a heuristic strategy that introduces a new partition whenever the call is non-recursive or the calling statement is definitely executed for every possible execution of the model, and a common partition representing the called method otherwise. The strategy can be further enhanced by bounding the depth of the call stack, possibly with different bounds for different categories of methods. The analysis and identification of useful heuristics constitutes an important part of future model refinement. To keep the analysis feasible, sparse representation of environments have to be used (see Section 6 for more details).

Another key feature in the analysis of object-oriented programs is the abstraction of heap data structures and aliased variables. For abstracting heap data structures, any of the numerous heap abstraction approaches developed in the last decades can be applied. For simplicity, we propose the approach given by [Cor98], where objects are abstracted into equivalence classes associated with the program point at which they were created. Note that simple approaches can lead to accurate results, as the partitioning strategies, together with the information from test cases, in many cases eliminate the need for approximation.

4.2 Analyzing Loops

For simplicity, we restrict the following discussion to `while` loops (other forms of iteration statements are treated similarly).

In case the condition of the loop can be evaluated uniquely using the abstract environment, the corresponding branch is followed, unless a termination check is triggered. Otherwise, conventional static analysis of the loop is done. The environment before and after the loop, together with assertions from the test case, are used as entry and exit states for the analysis. Static analysis is used to strengthen the pre- and postconditions of the loop, which are subsequently used to derive conflicts.

Note that in this approach, more precise results than with static analysis alone can be derived. This is because the values in the pre- and post environment are derived from test cases and, therefore, may be more precise than a purely static approximation. These values can in turn be used to derive approximation operators for the static analysis, which are effective at proving a contradiction. This idea is similar to [CDH⁺00], where abstraction operators are guessed based on the structure of the program and the given proof obligation.

Nontermination,³ by necessity, has to be dealt with through heuristics, such as setting upper limits on loop iterations or recursion depth, or using the abstract interpretation model to determine

²Here, the assumption is made that the program is terminating; this issue is revisited below.

³Although we assume the given program is terminating on all test cases, nonterminating programs can arise due to abnormality assumptions set by the diagnostic engine.

if any of the successor statements of the loop (call) can be reached using the current environment as entry state for the loop (call). This is an avenue for future research.

4.3 Correcting Faults

Once the possible fault locations have been narrowed down to a single candidate, heuristic algorithms to guess a replacement for the faulty instruction are applied. From the values of the environment before and after the faulty statement, and from the statements internal structure, instructions are synthesized to replace the incorrect statement [SW99]. This process can be aided by complementary models (Section 4.4), to restrict the search space for candidate instructions. We briefly sketch the synthesis algorithm; see [SW99] for further details.

Let s be the statement to be replaced. The set of replacement statements for s is derived according to the type of s and is constrained by a parameter k that limits the maximum size of the replacement. Replacements that do not satisfy the static type declarations in the original program are not considered, as we are only interested in corrections that satisfy the basic requirements of the language and the compiler. Possible replacements for constants and variables are constants, different variables, or calls to methods defined in the program. Method calls can be replaced either by a call to a different method (using a subset of the original calls arguments, or synthesizing new arguments), or by one of it's arguments (if any). For each replacement, a penalty measure ("size") is assigned, which measures the deviation from the original program, with zero being no modification.

The algorithm finds suitable replacements by enumerating possible replacements up to size k , ignoring all candidates that are inconsistent with the types and values derived for the diagnosis candidate $ab(s)$. Thus replacements more similar to the original program are tested earlier and are preferred to less similar statements. This algorithm can be extended to incorporate information provided by complementary models, as is demonstrated in the example in Section 5. In this case, variables that are indicated by the complementary model are assigned lower penalty values than other variables, resulting in former candidates being preferred.

For example, statement `float circ = 2.f*r*3.141f` has the following replacement candidates: replacing either constant with another constant (size 1) or with a variable (size 2), or with a method call without arguments (size 2). `r` can be replaced by a constant, another variable, or a method call without arguments (size 1, 1, and 2, respectively). Operators can be replaced with any of their arguments, a different operator, or a method call taking two arguments. Finally, the variable on the left hand side of the assignment can be replaced with any other assignable variable of the same type.

The modification of the model to incorporate the replacement instructions is done by introducing specialized mode assumptions $syn(\underline{l},r)$ and $\neg syn(\underline{l})$, where l is the program point where the replacement r is applied. $\neg syn(\underline{l})$ expressed that no modification takes place at l .

4.4 Complementary Models

Past experiences with MBD have shown that MBD provides excellent results when diagnosing functional faults. However, for structural faults, the semantics-based models do not provide sufficient information to accurately detect such faults. Furthermore, unless the test case specification is unreasonably detailed, for many programs a large number of diagnoses remains.

To overcome these problems, [Stu01] proposed to utilize complementary models, in particular representations of design information, to obtain the necessary information to guide the diagnostic engine. An advantage of this approach is that it integrates nicely into our framework without placing additional burden on the user.

For example, consider the state diagram in Figure 2. The automaton can be interpreted as a specification expressing that for each object the first call to method `getValue` (if any) must be preceded by a call to `setValue`. Translated into assertion statements and incorporated into our debugging engine, this model can be used to detect wrong or missing method calls, as demonstrated in the next section.

```
1  class Item {
2    int value;
3    void setValue(int v) { value = v; }
4    int getValue() { return value; }
5  }
6  class Main {
7    Item[] items;
8    int first, last;
9    /** @pre: (n > 0) */
10   void setup(int n, int d) {
11     int i = 0;
12     int k = 1;
13     items = new Item[n];
14     while (i < items.length) {
15       Item item = new Item();
16       items[i] = item;
17       k *= d;
18       i++;
19     }
20     first = items[0].getValue();
21     last = items[n-1].getValue();
22   }
23   /** @post: (first == d) &&
24    * (last == Math.pow(d,n)) &&
25    * (items.length == n) */
26 }
```

Figure 1: Example Program

Specifically, we propose to use partial specifications, i.e. pre- and postconditions and assertions, to generate additional conflicts. This provides multiple advantages:

- Paths of the model can be eliminated and possibly new conflicts be generated by checking the assertions.
- Assertions can be valid for multiple test cases, which avoids specifying program behavior separately for every test case and values.
- By comparing the dependencies between variables in pre- and postconditions, structural faults, such as wrong assignments or missing statements, can be detected [Stu01]. A similar approach, but without exploiting test case information, was used in [Jac95].

5 Example

This section puts together all the previous sections and demonstrates the framework's ability to locate and correct faults.

In this example, the interval abstraction from [CC77] is used to approximate a set of integer values. The model is structured such that diagnosis components represent single statements. For simplicity, hierarchic modeling is not applied, as the method's structure is rather simple. Also, the example does not require termination heuristics for loops or method calls. For objects created on the

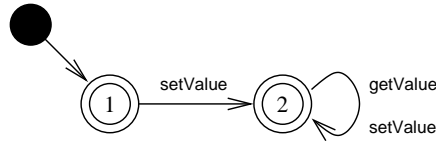


Figure 2: Valid Call Sequences for Item

heap, a simple abstraction aggregating all objects created at a specific location into one abstract model variable, corresponding to the allocation site, is used [Cor98].

Consider the program in Figure 1, where the statement `item.setValue(k)` is missing after line 17. Further, assume a test case $T = \{\llbracket 10 \rrbracket \mapsto \langle n = 3, d = 2 \rangle, \llbracket 22 \rrbracket \mapsto \langle items[k].value = 2^{k+1} (k \in [0..2]) \rangle\}$ ⁴ and the contract specification given in the method comments. Note that the test case *specification* expresses the *intended* result of the program. Also, we are given a complementary model that specifies that for each instance of class `Item`, the method `setValue()` must be called before `getValue()` is invoked (see Figure 2). This is translated into a separate instance variable `Item._callstate`, which is initialized to 1, denoting the state before the first method call in Figure 2. At the entry point of method `setValue()` `_callstate=2` is inserted, indicating that after the method is called, the automaton in Figure 2 is in state 2. Similarly, in method `getValue()`, the assertion always `_callstate==2` is inserted.⁵

When the method `setup()` is analyzed using the test case, the conflict $\mathcal{C} = \{\neg ab(\llbracket 13 \rrbracket), \neg ab(\llbracket 15 \rrbracket), \neg ab(\llbracket 20 \rrbracket)\}$ is derived: $\llbracket 12 \rrbracket$, $\llbracket 17 \rrbracket$, and $\llbracket 3 \rrbracket$ do not influence the result or the call sequence at all and can therefore be removed; when $\llbracket 4 \rrbracket$, $\llbracket 11 \rrbracket$, $\llbracket 14 \rrbracket$, $\llbracket 16 \rrbracket$, and $\llbracket 18 \rrbracket$ are abnormal, the complementary model in Figure 2 still derives a conflict with $\llbracket 20 \rrbracket$, as `setValue()` is not called before `getValue()`; $\llbracket 21 \rrbracket$ can be removed for the same reason.

The diagnostic process continues with the assumption that at least one component in \mathcal{C} is faulty. Rerunning the model for each component, the following conflicts are derived: $ab(\llbracket 13 \rrbracket)$ conflicts with $\{\neg ab(\llbracket 4 \rrbracket), \neg ab(\llbracket 11 \rrbracket), \neg ab(\llbracket 14 \rrbracket), \neg ab(\llbracket 15 \rrbracket), \neg ab(\llbracket 16 \rrbracket), \neg ab(\llbracket 18 \rrbracket), \neg ab(\llbracket 20 \rrbracket)\}$ because replacing `new Item[n]` with another expression still causes a contradiction for `first` in line 20, or a `NullPointerException`. For $ab(\llbracket 15 \rrbracket)$, no type compatible replacement for `new Item()` exists and, therefore, this assumption is not consistent either. $ab(\llbracket 20 \rrbracket)$ induces the conflict $\{\neg ab(\llbracket 13 \rrbracket), \neg ab(\llbracket 15 \rrbracket), \neg ab(\llbracket 16 \rrbracket), \neg ab(\llbracket 21 \rrbracket)\}$, as $\llbracket 21 \rrbracket$ either causes a `NullPointerException`, or the complementary model again derives a call sequence conflict.

As none of the attempts to restore consistency by assuming abnormality for any of the components of the initial conflict is successful, no single-fault diagnosis exists for the given program and test case. As a consequence, the diagnostic process has to choose between increasing the diagnosis cardinality or to search for structural faults. As we are interested in simple faults, it is reasonable to look for structural faults before increasing diagnosis cardinality.

Using the model from Figure 2, and from the conflicts above, it can be deduced that a call to `setValue()` is missing. From the conflict \mathcal{C} and the knowledge that the objects causing a contradiction were created in $\llbracket 15 \rrbracket$, the missing call can be inserted between $\llbracket 15 \rrbracket$ and $\llbracket 20 \rrbracket$ (denoted $\llbracket 15' \rrbracket$ – $\llbracket 19' \rrbracket$).

The diagnostic process is restarted, using four new rules to insert a synthesized statement s' (Section 4.3) at location l whenever $syn(\llbracket l \rrbracket, s')$ is assumed (with $\neg syn(\llbracket l \rrbracket)$ being the default). The simplest candidates including a call to `setValue()` are of the form $\alpha.setValue(\beta)$, where $\alpha \in \{item, items[\alpha']\}$, $\alpha', \beta \in \{first, last, i, k, n, d\}$.

To further restrict the synthesis candidates, we utilize dependency information provided by a complementary model: The postconditions of the method imply a dependency from variable `d` to the variables `first` and `last`. On the other hand, these dependencies cannot be derived from the implementation. As `first` and `last` depend on `items[.].value` only, either `first` and `last` directly, or `items[.].value` must also depend on `d`. Therefore, the synthesized statements using `d` or `k` as argument to `setValue()`

⁴Components corresponding to statements are identified by the statement's line number.

⁵Note that the mapping from sequence diagrams to assertions is more complicated if a state diagram contains multiple paths leading to a method. For simplicity, we refrain from a detailed discussion in this paper.

are preferred. This example also illustrates that it can be advantageous to express assertions about programs in terms of variables of the program instead of test case specific values, where comparing dependencies is not possible.

With models that have been modified by adding the synthesized expressions, three diagnoses are obtained: $syn(\boxed{15'}, \text{items}[i].\text{setValue}(k))$ and $syn(\boxed{l}, \text{item}.\text{setValue}(k))$ with $l \in \{17', 18'\}$. Other candidates are not consistent for the following reasons:

Location	Candidate	Conflict
$\boxed{15'} - \boxed{18'}$	$\text{item}.\text{setValue}(d)$	Contradiction with test case (for first or last)
$\boxed{15'}$, $\boxed{16'}$	$\text{item}.\text{setValue}(k)$	Contradiction with test case (for first or last)
$\boxed{17'}$, $\boxed{18'}$	$\text{item}.\text{setValue}(k)$	—
$\boxed{15'}$	$\text{items}[\alpha].\text{setValue}(\beta)$	Uncaught NullPointerException
$\boxed{16'} - \boxed{19'}$	$\text{items}[\gamma].\text{setValue}(\beta)$	Contradiction with test case (for first or last)
$\boxed{16'} - \boxed{19'}$	$\text{items}[k].\text{setValue}(\beta)$	Uncaught ArrayIndexOutOfBoundsException
$\boxed{18'}$, $\boxed{19'}$	$\text{items}[i].\text{setValue}(\beta)$	Uncaught ArrayIndexOutOfBoundsException
$\boxed{16'}$	$\text{items}[i].\text{setValue}(\beta)$	Contradiction with test case (for first or last)
$\boxed{17'}$	$\text{items}[i].\text{setValue}(d)$	Contradiction with test case (for first or last)
$\boxed{17'}$	$\text{items}[i].\text{setValue}(k)$	—

$(\alpha \in \{d, k, n, \text{first}, \text{last}\}, \beta \in \{d, k\}, \gamma \in \{d, n, \text{first}, \text{last}\})$

Note that the combination of dynamic execution and static analysis is more powerful than static analysis alone. As a demonstration, consider the synthesized statement $\text{items}[d].\text{setValue}(d)$. Our model is able to derive a conflict with variable first , whereas static analysis cannot because elements of the array are approximated as $[0..d]$ (assuming an interval abstraction [CC77] for Item.value , and a heap abstraction that does not distinguish between $\text{items}[0]$ and $\text{items}[d]$).

Finally, the suggestion to insert either $\text{items}[i].\text{setValue}(k)$ after line 17, or $\text{item}.\text{setValue}(k)$ after lines 17 or 18 is presented to the user.

6 Sparse Trace Representation

This section describes a generalized notion of program trace and a sparse representation thereof. This representation makes it possible to avoid copying parts of the dynamic data structures created by a program, as was required by previous models [May00].

Definition 4 (Variable Identifier) *A variable identifier is either the canonical name of a local or static variable, or is composed of an object identifier o and a canonical name of an instance variable v (denoted $o.v$).*

The canonical name of a variable is formed by prefixing its name with the fully qualified name of the scope the variable is defined in. For example, the canonical name of static variable out defined in class System , which is defined in package java.lang , is $\text{java.lang.System.out}$. Object identifiers are an abstraction of memory addresses for objects created on the heap. In this work, we use the statement that created the object as identifier. For multi-dimensional arrays, the index of the parent array is included to distinguish different sub-arrays.

Definition 5 (Environment) *An environment e is a tuple $\langle c : l, \mathcal{V} \rangle$, where $c \in \mathcal{C}$ is a unique context identifier and $l \in \mathcal{L}$ the label of the statement e is associated with. \mathcal{V} is a mapping from variable identifiers into abstract values. \mathcal{E} denotes the set of all abstract environments \mathcal{V} .*

In this work, context and partition are used in the sense of points-to analysis or call graph construction [GDDC97] and represent an abstraction of the call site of a method. Context identifiers are used to distinguish different instances of a program part during analysis. For example, if a method is called multiple times in a trace, the analysis of the two calls can be merged into a single analysis, using the merged input and output environments of both calls. While speeding up the analysis,

merging, i.e. partitioning, contexts results in diminished accuracy and is applied only when necessary. Traces without loops can be analyzed without merging, while recursive calls or loop statements may require approximation in case the recursion depth or the number of iterations cannot be determined.

The relationship between concrete and abstract values is given by the abstraction function selected for the abstract interpretation. An abstract environment associates each variable identifier with an abstract value, thus approximating the set of concrete values in a non-relational way. For example, the well-known interval abstraction [CC77] approximates a set of integer values with an interval spanning all the values in the set.

Every program is transformed into a simple intermediate representation, consisting of assignments, primitive operations and method calls at top-level.

Definition 6 (Program) A program P is a pair $\langle S, \mathcal{R} \rangle$, where S is a finite set of statements $l : s$, each labeled with a unique label $l \in \mathcal{L}$. $\mathcal{R} \subseteq \Gamma(\mathcal{E}) \times \Gamma(\mathcal{E})$ is a transfer relation, specifying the possible transitions between concrete environments. $\Gamma(\mathcal{E}) \stackrel{\text{def}}{=} \{\gamma(e) | e \in \mathcal{E}\}$.

For short, $\langle a, b \rangle \in \mathcal{R}$ is denoted $a \xrightarrow{\mathcal{R}} b$.

A context selection function C generates a label for the destination environment, given an environment.

Definition 7 (Execution Trace) The execution trace \mathcal{T} of P is defined as $\mathcal{T} = \bigcup_{i \geq 0} \mathcal{T}^i$, with $\mathcal{T}^0 = \{\exists_p \gamma(e_0) \xrightarrow{\mathcal{R}} p\}$, $\mathcal{T}^{i+1} = \mathcal{T}^i \cup \{p \rightarrow q | \exists_r r \xrightarrow{\mathcal{T}^i} p, \gamma(p) \xrightarrow{\mathcal{R}} \gamma(s), s = \langle c : l, \mathcal{V} \rangle, c' = C(s), q = \langle c' : l, \text{map}(\mathcal{T}^i, c' : l) \sqcup \mathcal{V} \rangle\}$. e_0 denotes the abstract environment at the starting point of the trace. $\text{map}(\mathcal{T}^i, c' : l)$ denotes the variable mapping for the environment labeled $c' : l$ in \mathcal{T}^i (\perp if none exists), and \sqcup is the join operator of the abstract domain lattice.

This definition builds the graph containing all feasible paths, starting from e_0 . Given the reachable environments from the previous iteration, new transitions are added leading to the reachable environments as specified by \mathcal{R} .

The context selection function $C : \mathcal{E} \mapsto \mathcal{C}$ determines the context of the target environment, given the source environment. For recursive method calls and loops, infinite execution sequences have to be finitely approximated by partitioning the set of all execution contexts into finitely many partitions. C can be influenced by the user or by heuristics to adjust the degree of imprecision. As mentioned in Section 4.2, a bounded approximation of the call stack and loop counter can be used to analyze recursive and iterative program constructs. For other program elements, $C = Id$ is sufficient, as no approximation is necessary.

Building on the (Static Single Information form (SSI) [Ana99, MS02], a separate copy of all used variables is created for each branch in the execution trace. The SSI form makes use of ϕ and σ functions, which deal with data flow information spanning control flow paths. ϕ functions are placed at control flow join points and combine the values from the incoming branches into an approximate value that is used for references to that variable below the ϕ function. Similarly, σ functions are placed at control flow branches and create separate copies of the incoming value for each branch. Instead of computing locations for σ and ϕ functions statically, we utilize test case information to restrict execution paths. This makes it possible to omit many of the σ and ϕ functions, which in turn produces simpler models.

6.1 Incremental Trace Construction

To allow for efficient computation of the used and modified variables, the trace representation is split into two parts. One part is a static approximation of all possible traces, used to determine basic blocks where ϕ and σ functions may be necessary. The other part represents the feasible execution paths through the program, starting from e_0 .

By constructing the trace incrementally while executing a test case, the used and modified variables, and in particular, the used and modified objects, are known. As a consequence, only the objects that are actually modified by a statement need to be updated, leaving all other objects unmodified. Consequently, subsequent statements are connected directly to the last modification of their used objects, instead of being connected to the last modification of *any* instance of the same type. Therefore, copying values of unmodified instances is unnecessary and can be omitted.

The static approximation \mathcal{T}_S of $C \circ \mathcal{R}$ for each method is derived from a control flow graph (CFG) as described in [MS02]. This is possible as Section 4.1 restricts C such that only for method entry nodes or loop headers, $C \neq Id$ is possible. From the approximation, the *DF graph* [Sre95] is constructed. The DF graph contains all nodes of the trace's CFG, indicating nodes which are locations of possible ϕ functions. From each node n in the graph, links point to the ϕ functions in the dominance frontier of n .⁶ Therefore, by reversing the direction of the links, each ϕ function is linked to the collection of nodes which give rise to ϕ . Each node is labeled with a unique identifier.

The representation \mathcal{T}_D of all feasible execution paths is built by repeatedly applying Definition 7, starting from the entry environment e_0 . Transitions are grouped into basic blocks, with linear transition sequences⁷ being compressed into one block.

Environments with multiple outgoing transitions in \mathcal{T}_D denote the end of the current basic block. For each outgoing transition $p \rightarrow q$ with $q \notin \mathcal{T}_D$, new basic blocks are created. Otherwise, q already exists and just gains a new incoming link. At this point, several steps are necessary to preserve the correctness of \mathcal{T}_D :

First, in case q is not the first environment of a block B , B is split into two parts, B_0 and B_1 , consisting of the transitions leading to q and the remaining path starting at q , respectively. B_1 is inserted as a successor of B_0 and the link $p \rightarrow q$ is added.

Next, the set $\mathcal{B} = DF^{-1}(q)$ of blocks giving rise to a ϕ function at this environment q is determined using the links in \mathcal{T}_S , where only blocks are considered that are actually instantiated in \mathcal{T}_D for the current context. For each $b \in \mathcal{B}$, the set of modified variables is determined and ϕ functions are created for each variable (unless they already exist).

To maintain correctness of \mathcal{T}_D , the ordering in which the transitions are processed is crucial. It must be ensured that all modified variables for a block are known before the block is used to generate other ϕ functions. This can be ensured by suspending the processing of ϕ function generation, in case not all blocks of the current context corresponding to blocks in \mathcal{B} have been analyzed completely, or are known to be unreachable. In addition, an ordering has to be imposed on contexts and labels, such that loops and called methods are analyzed completely before any of the successor transitions are expanded. If assuming the proposed context selection strategy from Section 4.1, this is not a severe restriction for our framework.

If the graph is cyclic, this ordering is not enough, and a fixpoint algorithm needs to be used (details are omitted for brevity).

The introduction of σ functions [Ana99] for used variables is handled similarly to ϕ functions. Possible locations for σ functions are computed using \mathcal{T}_S with the direction of all arcs reversed, and an auxiliary environment that postdominates⁸ all exit environments in the original \mathcal{T}_S . Special treatment of cyclic structures is not necessary in this case.

Whenever a branch of the trace is found inconsistent, the branch is removed and replaced with a summary of the part of the derivation of the inconsistency that is local to the branch. In case the branch is the only outgoing or incoming connection to a σ or ϕ function, the function is removed and all used associated variable are redirected to the previous definition. Although branch elimination is not necessary for the initial forward trace construction (as only consistent branches are followed), existing inconsistent branches may be found in subsequent backward and forward iterations.

⁶The dominance frontier of a node n contains all nodes n' of which n dominates an immediate predecessor of n' , but not n' itself. It can be shown that these are exactly the locations where ϕ functions need to be placed.

⁷A linear transition sequence is a nonempty sequence $\langle e_i \rightarrow e_{i+1} \rightarrow \dots \rightarrow e_{i+k} \rangle$, where none of the $e_{i+1}, \dots, e_{i+k-1}$ has more than one predecessor or successor.

⁸An environment e_1 postdominates e_2 iff all paths from e_2 to the exit environment visit e_1 .

6.2 Complexity

The time complexity of the trace construction is $\mathcal{O}(n_D \cdot \max(n_S, n_D)^2 \cdot \alpha(n_D))$ in the worst case, where n_S and n_D denote the number of blocks in \mathcal{T}_S and the number of environments in \mathcal{T}_D , respectively. $\alpha(n_D)$ represents the worst case complexity of the fixpoint computation, which depends on the program structure and on the abstract domain lattice.

7 Related Work

Automated debugging has been an active area of research for several decades, resulting in a large number of different methodologies using various assumptions and algorithms.

In Program Slicing [Wei84, Tip95], statements that cannot influence the value of a variable at a given program point are eliminated by considering the dependencies between the statements. Backward reasoning from output values, as in our approach, is not possible. Similar ideas were successfully utilized in a MBD tool analyzing VHDL programs [FSW99, Wot01].

[BH93, BH95] use probability measurements to guide diagnosis. The program debugging process is divided into two steps. In the first one, program parts that may cause a discrepancy are computed by tracing the incorrect output back to the inputs and collecting the involved statements. In a second step, a belief network is used to identify the most probable statements causing the fault. Although this approach was successful in debugging a very large program, it requires statistics relating the statement types and fault symptoms, which makes it unsuitable for debugging general programs.

The idea of path information to guide debugging was also applied by other researchers, such as program dicing [Tip95] and similar heuristics [PS92] and visualization of test results [JHS02]. Whereas those ideas seem to provide good results, they are even more valuable when integrated into a model-based debugging environment, as they can provide the necessary information to discriminate between diagnoses and aid the selection of more likely candidates [MSWW02b].

Jackson [Jac95] introduces a framework to detect faults in programs that manifest through changed dependencies between the input and the output variables of a program. The approach detects differences between the dependencies computed for a program and the dependencies specified by the user. It is able to detect certain kinds of structural faults but no test case information is exploited. Whereas Jackson focuses on bug detection, the model-based approach is also capable of locating faults. Further, the information obtained from present and absent dependencies can aid the debugger to focus on certain regions and types of faults, and thus find possible causes more quickly.

[Hun98] applies the idea of MBD to the domain of object-oriented languages by building models for programs written in Smalltalk. The model used in his work is based on dependencies between instance variables and method calls that modify them. The observations state whether the computed value of a variable is correct or not, regardless of its concrete value. This approach is limited to programs that contain a single faulty statement. Also, previous results showed [Wie01] that while dependencies are a valuable tool to isolate faulty modules, more expressive models are needed to locate faults on a finer-grained level and to reduce frequent user-interaction.

[HZ00] introduces an algorithm that compares a faulty program to a close correct variant to determine changes that cause the misbehavior. Although the algorithm seems to be highly effective for test case minimization and has also been applied to locate failure causes in programs [CZ00], the approach generally requires a close and correct variant of the program (or a preselection of “interesting” statements, for example in form of grouped changes from a versioning system) to be effective.

[CFD93] were the first to study model-based debugging, with logic programs as language of interest. Their approach was later extended and refined by [Bon94, BP94]. Their approach connects diagnosis and debugging by identifying horn clauses to be added or removed from programs to fix a fault. They show that the MBD approach is more efficient in terms of user interaction than Algorithmic Debugging [Sha83].

[FFJS00] apply similar ideas to knowledge base maintenance, exploiting hierarchical information to speed up the diagnostic process and to reduce the number of diagnoses.

Following [CFD93], MBD was extended to imperative and concurrent languages, in particular to a subset of VHDL [FSW99]. This work showed that MBD can be successfully applied in this domain to isolate faulty processes. Diagnosing programs at a finer level of granularity is still ongoing research [PW03] and requires overcoming difficulties related to temporal and concurrency-related aspects of the VHDL language.

Mateis et al. [MSW00] introduce a dependency-based model for Java programs that abstracts from concrete variable values. However, for programs with complex structure, either a high amount of user-provided information is necessary, or the results are relatively coarse. In [May00] it was extended to simulate program execution. The models are limited to structured, non-recursive programs and are not as expressive as the abstract-interpretation-based approach when the behavior of complex components is only partially deducible given a test case and diagnostic assumptions.

Previous research in MBD has resulted in a set of tools that successfully demonstrated the potential of the approach. The main strength of the model-based techniques is that reasoning strategies are separated from conflict detection, which makes it feasible to plug-in a variety of program analysis and debugging methods, provided the results of the analysis can be mapped back to the program's source code. A number of models have been developed and analyzed, resulting in promising results, mainly in the domain of functional faults (such as wrong constants, operators, conditional expressions, etc.). However, the combination of multiple models and reasoning strategies to improve accuracy and reduce user interaction is still ongoing research and needs further evaluation, in particular with a larger set of realistic programs. This work aims at making a first step in this direction by combining abstract-interpretation-based models with complementary models to correct omitted statements and structural faults. Also, the implementation of most of the models currently is only incomplete and experimental. In particular, no optimizations for speed have been done, which makes the comparison with other approaches rather difficult.

Abstract Interpretation to analyze programs was first introduced by [CC77], and later extended by [Bou93, CC00] to include assertions for abstract debugging. Their approach aims at analyzing every possible execution of a program, which makes it suitable to detect errors even in the case where no test cases are available. A common problem of these approaches is that of choosing appropriate abstractions in order to obtain useful results, which hinders the automatic applicability of these approaches for many programs. [Bou92] introduces a relaxed form of representation for abstract interpretation, which allows for more complex domains, while building the structure of the approximation dynamically. Our framework is strongly inspired by this work, but provides more insight on how to choose approximation operators for debugging, in particular in the case where test information is known. These questions are not addressed in [Bou92].

Recently, model checking approaches have been extended to attempt fault localization in counterexample traces. [BNR03] extended a model checking algorithm that is able to pinpoint transitions in traces responsible for a faulty behavior. [GV03] presents another approach, which explores the neighborhood of counterexamples to determine causes of faulty behavior. These techniques mostly consider deviations in control flow and do not take data dependencies into account. Also, the derivation of the abstract model from the concrete program usually is non-trivial and is difficult to automate.

8 Conclusion

We have presented an automatic debugging approach utilizing model-based diagnosis together with an abstract interpretation based conflict detection framework. Based on experiences with previous models [MSWW02a], this framework is able to detect large classes of programming errors, such as faulty expressions and faults in control flow, given a set of test cases and partial specifications of the programs behavior. This work extends the approach to provide more accurate results in cases where previous models could not derive conflicts by approximating loops and recursive function calls using abstract interpretation. Further, the introduction of complementary models allows to extend this ap-

proach to structural faults. The abstract interpretation framework makes it possible to parameterize the framework in various directions: the approximation of variable values can be chosen, heuristics for partitioning of context for static analysis and heap analysis are parameterizable, and heuristics for detection of nontermination are incorporated to avoid nonterminating diagnoses. The framework's ability to locate and correct certain faults automatically was demonstrated using a simple example program. Possible extensions are the representation for abstract domains from [Bou92], and the analysis and refinement of heuristics for context partitioning and termination detection. While those heuristics are not essential for our approach, abstractions tailored to specific programs and specifications [CDH⁺00] can improve the results dramatically.

References

- [Ana99] Scott Ananian. The static single information form. Master's thesis, Department of Electrical and Computer Science, Princeton University, 1999.
- [BH93] Lisa Burnell and Eric Horvitz. A synthesis of logical and probabilistic reasoning for program understanding and debugging. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, pages 285–291, 1993.
- [BH95] Lisa Burnell and Eric Horvitz. Structure and chance: Melding logic and probability for software debugging. *Communications of the ACM*, 38(3):31–41, 1995.
- [BNR03] Thomas Ball, Mayur Naik, and Sriram K. Rajamani. From symptom to cause: Localizing errors in counterexample traces. In *Proc. Symposium on Principles of Programming Languages*, 2003.
- [Bon94] Gregory W. Bond. *Logic Programs for Consistency-Based Diagnosis*. PhD thesis, Carleton University, Faculty of Engineering, Ottawa, Canada, 1994.
- [Bou92] François Bourdoncle. Abstract interpretation by dynamic partitioning. *Journal of Functional Programming*, 1992.
- [Bou93] François Bourdoncle. Abstract debugging of higher-order imperative languages. In *Proceedings of the SIGPLAN Conference on Programming Language Design and Implementation*, pages 46–55, 1993.
- [BP94] G. W. Bond and B. Pagurek. A Critical Analysis of “Model-Based Diagnosis Meets Error Diagnosis in Logic Programs”. Technical Report SCE-94-15, Carleton University, Dept. of Systems and Computer Engineering, Ottawa, Canada, 1994.
- [CC77] Patrick Cousot and Radhia Cousot. Abstract interpretation: A unified lattice model for static analysis of programs by construction of approximation of fixpoints. In *Proc. Symposium on Principles of Programming Languages*, pages 238–252, Los Angeles, California, January 1977.
- [CC00] Patrick Cousot and Radhia Cousot. Abstract interpretation based program testing. In *Proceedings of the SSGRR 2000 Computer & eBusiness International Conference*, 2000.
- [CDH⁺00] James Corbett, Matthew Dwyer, John Hatcliff, Corina Pasareanu, Robby, Shawn Laubach, and Hongjun Zheng. Bandera: Extracting finite-state models from Java source code. In *Proceedings of the 22nd International Conference on Software Engineering*, 2000.
- [CFD93] Luca Console, Gerhard Friedrich, and Daniele Theseider Dupré. Model-based diagnosis meets error diagnosis in logic programs. In *Proceedings 13th International Joint Conf. on Artificial Intelligence*, pages 1494–1499, Chambery, August 1993.

- [Col97] Christopher Colby. Accumulated imprecision in abstract interpretation. In *AAS'97*, 1997.
- [Cor98] James C. Corbett. Using shape analysis to reduce finite-state models of concurrent Java programs. Technical report, Department of Information and Computer Science, University of Hawaii, 1998.
- [CZ00] Holger Cleve and Andreas Zeller. Finding failure causes through automated testing. In Mireille Ducassé, editor, *Proceedings of the 4th International Workshop on Automated and Algorithmic Debugging, AADEBUG '00*, Munich, Germany, 2000.
- [FFJS00] Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach, and Markus Stumptner. Exploiting structural abstractions for consistency based diagnosis of large configurator knowledge bases. In *ECAI Workshop on Configuration*, Berlin, August 2000.
- [FSW99] Gerhard Friedrich, Markus Stumptner, and Franz Wotawa. Model-based diagnosis of hardware designs. *Artificial Intelligence*, 111(2):3–39, July 1999.
- [GDDC97] David Grove, Greg DeFouw, Jeffrey Dean, and Craig Chambers. Call graph construction in object-oriented languages. In *ACM Proceedings of the Conference on Object-Oriented Programming Systems, Languages, and Applications*, Atlanta, GA, October 1997.
- [GV03] Alex Groce and Willem Visser. What went wrong: Explaining counterexamples. In *SPIN Workshop on Model Checking of Software*, 2003.
- [HP00] Michael Hind and Anthony Pioli. Which pointer analysis should I use? In *Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*, 2000.
- [Hun98] John Hunt. Model-Based Software Diagnosis. *Applied Artificial Intelligence*, 12(4):289–308, 1998.
- [HZ00] Ralf Hildebrandt and Andreas Zeller. Simplifying failure-inducing input. In *Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*, Portland, OR, USA., August 2000.
- [Jac95] Daniel Jackson. Aspect: Detecting Bugs with Abstract Dependences. *ACM Transactions on Software Engineering and Methodology*, 4(2):109–145, April 1995.
- [JHS02] James A. Jones, Mary Jean Harrold, and John Stasko. Visualization of test information to assist fault localization. In *Proceedings of the 24th International Conference on Software Engineering*, Zurich, Switzerland, September 2002.
- [Llo87] J. W. Lloyd. Declarative Error Diagnosis. *New Generation Computing*, 5:133–154, 1987.
- [May00] Wolfgang Mayer. Modellbasierte Diagnose von Java-Programmen, Entwurf und Implementierung eines wertbasierten Modells. Master's thesis, Technische Universität Wien, Institut für Informationssysteme, 2000. (in German).
- [MS02] Wolfgang Mayer and Markus Stumptner. Modeling programs with unstructured control flow for debugging. In *Proc. 15th Australian Joint Conf. on AI*, pages 107–118, Canberra, December 2002. Springer-Verlag.
- [MSW00] Cristinel Mateis, Markus Stumptner, and Franz Wotawa. Modeling Java Programs for Diagnosis. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, Berlin, Germany, August 2000.

- [MSWW02a] Wolfgang Mayer, Markus Stumptner, Dominik Wieland, and Franz Wotawa. Can AI help to improve debugging substantially? Debugging Experiences with Value-Based Models. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 417–421, Lyon, 2002.
- [MSWW02b] Wolfgang Mayer, Markus Stumptner, Dominik Wieland, and Franz Wotawa. Towards an Integrated Debugging Environment. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, pages 422–426, Lyon, 2002.
- [MT02] Jakob Mauss and Mugur Tatar. Computing minimal conflicts for rich constraint languages. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, Lyon, 2002.
- [PS92] Hsin Pan and Eugene H. Spafford. Heuristics for automatic localization of software faults. Technical report, Purdue University, 1992.
- [PW03] Bernhard Peischl and Franz Wotawa. Towards a framework for automated debugging: Abstracting the temporal behavior of VHDL-RTL programs. In *Proceedings of the Fourteenth International Workshop on Principles of Diagnosis*, Washington, D.C., June 2003.
- [Rei87] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.
- [Sha83] Ehud Shapiro. *Algorithmic Program Debugging*. MIT Press, Cambridge, Massachusetts, 1983.
- [Sre95] Vugranam C. Sreedhar. *Efficient Program Analysis using DJ Graphs*. PhD thesis, School of Computer Science, McGill University, Montréal, 1995.
- [Stu01] Markus Stumptner. Using design information to identify structural software faults. In *Proc. 14th Australian Joint Conf. on AI*, Springer LNAI 2256, pages 473–486, Adelaide, December 2001.
- [SW99] Markus Stumptner and Franz Wotawa. Debugging Functional Programs. In *Proceedings 16th International Joint Conf. on Artificial Intelligence*, pages 1074–1079, Stockholm, Sweden, August 1999.
- [Tip95] Frank Tip. A Survey of Program Slicing Techniques. *Journal of Programming Languages*, 3(3):121–189, September 1995.
- [Wei84] Mark Weiser. Program slicing. *IEEE Transactions on Software Engineering*, 10(4):352–357, July 1984.
- [Wie01] Dominik Wieland. *Model-Based Debugging of Java Programs Using Dependencies*. PhD thesis, Technische Universität Wien, November 2001.
- [Wot01] Franz Wotawa. On the Relationship between Model-based Debugging and Programm Mutation. In *Proceedings of the Twelfth International Workshop on Principles of Diagnosis*, Sansicario, Italy, 2001.