The abc Group

A staged static program analysis to improve the performance of runtime monitoring

(extended version)

abc Technical Report No. abc-2007-2

Eric Bodden[1], Laurie Hendren[1], Ondřej Lhoták[2]

[1] McGill University, Montréal, Québec, Canada
[2] University of Waterloo, Waterloo, Ontario, Canada

April 25, 2007

aspectbench.org

Contents

List of Figures

Abstract

In runtime monitoring, a programmer specifies a piece of code to execute when a trace of events occurs during program execution. Our work is based on tracematches, an extension to AspectJ, which allows programmers to specify traces via regular expressions with free variables. In this paper we present a staged static analysis which speeds up trace matching by reducing the required runtime instrumentation.

The first stage is a simple analysis that rules out entire tracematches, just based on the names of symbols. In the second stage, a points-to analysis is used, along with a flow-insensitive analysis that eliminates instrumentation points with inconsistent variable bindings. In the third stage the points-to analysis is combined with a flow-sensitive analysis that also takes into consideration the order in which the symbols may execute.

To examine the effectiveness of each stage, we experimented with a set of nine tracematches applied to the DaCapo benchmark suite. We found that about 25% of the tracematch/benchmark combinations had instrumentation overheads greater than 10%. In these cases the first two stages work well for certain classes of tracematches, often leading to significant performance improvements. Somewhat surprisingly, we found the third, flow-sensitive, stage did not add any improvements.

1 Introduction

Various mechanisms have been proposed for monitoring programs as they run. Aspect-oriented programming (AOP) is one approach where a programmer specifies which events should be intercepted and what actions should be taken at those interception points. More recently, this concept of event matching has been further expanded to include matching of *traces of events* [1, 20, 25, 29]. While this expanded notion of matching on traces is much more powerful, it can also lead to larger runtime overheads since some information about the runtime history must be maintained in order to detect matching traces. Also, instrumentation needs to be put in place in order to update this information at events of interest.

In this paper, we examine the problem of improving runtime performance of *tracematches*. Tracematches are an extension to AspectJ which allows programmers to specify traces via regular expressions of symbols with free variables [1]. Those variables can bind objects at runtime, a crucial feature for reasoning about object-oriented programs. When a trace is matched by a tracematch, with consistent variable bindings, the action associated with the tracematch executes. Trace matching is implemented via a finite-state-based runtime monitor. Each event of the execution trace that matches a declared symbol in a tracematch causes the runtime monitor to update its internal state. When the monitor finds a consistent match for a trace, it executes its associated action.

There are two complementary approaches to reducing the overhead for this kind of runtime monitoring. The first line of attack is to optimize the monitor itself so that each update to the monitor is as inexpensive as possible and so that unnecessary state history is eliminated. Avgustinov et al. were able to show that these approaches greatly reduce overheads in many cases [5, 6]. However, as our experimental results show, there remain a number of cases where the overhead is still quite large.

Our work is the second line of attack, to be used when significant overheads remain. Our approach is based on analysis of both the tracematch specification and the whole program being monitored. The analysis determines which events do not need to be monitored, i.e. which instrumentation points can be eliminated. In the best case, we can determine that a tracematch never matches and all overhead can be removed. In other cases, our objective is to minimize the number of instrumentation points required, thus reducing the overhead.

In developing our analyses, we decided to take a staged approach, applying a sequence of analyses, starting with the simplest and fastest methods and progressing to more expensive and more precise analyses. An important aspect of our research is to determine if the later stages are worth implementing, or if the earlier stages can achieve most of the benefit. We have developed three stages where each stage adds precision to our abstraction. The first stage, called the *quick check*, is a simple method for ruling out entire tracematches, just using the *names* of symbols. Our second stage uses a demand-driven [23] points-to analysis [10, 14], along with a flow-insensitive analysis of the program, to eliminate instrumentation points with inconsistent *variable bindings*. The third stage combines the points-to analysis with a flow-sensitive analysis that takes into consideration the *order* in which events may occur during runtime.

We have evaluated our approach using the DaCapo benchmark suite [7] and a set of 9 tracematches. We found that even though previous techniques often kept the runtime overhead reasonable, there were a significant number of benchmark/tracematch combinations which led to a runtime overhead greater than 10%. We focused on these cases and found that our first two stages worked well for certain classes of tracematches. We were somewhat surprised to find that our third stage did not add any further accuracy, even though it was the only flow-sensitive analysis, and we provide some discussion of why this is so.

This paper is organized as follows. Section 2 introduces tracematches, explains how they apply to Java programs, and gives some examples of where monitoring instrumentation can statically be shown to be unnecessary. In Section 3 we present our staged static analysis which performs such detection automatically. We carefully evaluate our work in Section 4, showing which problem cases our analysis can handle well, but also which cases might need more work or will probably never be statically analyzable. In Section 5 we discuss related analyses, finally concluding in Section 6. There we also briefly discuss our intended scope for future work on the topic.

2 Background

A tracematch defines a runtime monitor using a declarative specification in the form of a regular expression. The alphabet of this regular expression consists of a set of symbols, where one defines each symbol via an AspectJ pointcut. A *pointcut* is, in general, a predicate over joinpoints, a *joinpoint* in AspectJ being an event in the program execution. Common pointcuts can be used to specify a pattern to match against the name of the currently executing method or against the currently executing type. Special pointcuts also allow one to expose parts of the execution context. For instance, in the original AspectJ language the programmer can bind the caller and callee objects as well as all call arguments for each method call. We however use our own implementation of AspectJ in form of the AspectBench Compiler (`abc`) [3], which implements tracematches and with respect to context exposure also allows one to access any objects that can be reached or computed from the objects one can bind in plain AspectJ, or from static members [5]. For more details regarding pointcuts in AspectJ, see [2].

An example is shown in Figure 1. This tracematch checks for illegal program executions where a vector is updated while an enumeration is iterating over the same vector. First, in lines 2-5 it defines a plain AspectJ pointcut capturing all possible ways in which a vector could be updated. The actual tracematch follows in lines 7-13. In its header (line 7) it declares that it will bind a Vector v and an Enumeration e. Then, in lines 8-10 it defines the alphabet of its regular expression by stating the symbols create, next and update. The first one, create, is declared to match whenever any enumeration e for v is created, while next matches when the program advances e and update on any modification of v.

```
1  aspect FailSafeEnum {
2    pointcut vector_update() :
3      call(* Vector.add*(..))  ||  call(* Vector.clear ())  ||
4      call(* Vector.insertElementAt(..))  ||  call(* Vector.remove*(..))  ||
5      call(* Vector.retainAll (..))  ||  call(* Vector.set *(..));
6
7    tracematch(Vector v, Enumeration e) {
8       sym create after returning(e) : call(* Vector+.elements()) && target(v);
9       sym next before : call(Object Enumeration.nextElement()) && target(e);
10      sym update after : vector_update() && target(v);
11
12     create next* update+ next { /* handle error */ }
13   }
14 }
```

Figure 1: Safe enumeration tracematch

Line 12 declares a regular expression that states when the tracematch body (also line 12) should execute.

This should be the case whenever an enumeration was created, then possibly advanced multiple times and then at least one update to the vector occurs, lastly followed by another call to Enumeration.nextElement().

The declarative semantics of tracematches state that the tracematch body should be executed for any sub-sequence of the program execution trace that is matched by the regular expression with a consistent variable binding. A variable binding is consistent when at every joinpoint in the sub-sequence each variable is bound to the same object.

Internally, each tracematch is implemented using a finite state machine. Such state machines are similar to state machines that can be used for verification of typestate properties [26]. In such a property, a state machine can be associated with a single object. Whenever certain methods on that object are invoked, this state machine is updated according to its transition table. If during the execution a special error state is reached, the typestate property is violated.

Tracematches can be seen as an implementation of checkers for *generalized* typestate properties [17]. While ordinary typestate properties only reason about a single object, the generalized ones allow reasoning about groups of objects. Consequently, the tracematch implementation needs to associate a state not with a single object but rather with a group of objects, stored as mapping from tracematch variables to Java objects. Due to their semantic foundations [1], those mappings are called *disjuncts*. Because multiple such groups of objects can be associated with the same automaton state at the same time, each state of the automaton is associated with a *set* of disjuncts, which we call a *constraint*. (Semantically, as shown in [1], this implementation represents storing object constraints in Disjunctive Normal Form.)

When compiling a program that contains a tracematch, the compiler firstly generates program code for the related state machine and secondly instruments the program such that it notifies the state machine about any joinpoint of interest, i.e. any joinpoint that matches any of the declared symbols of the tracematch. When such a notification occurs, the related state machine updates its internal state accordingly, i.e. propagates disjuncts from one state to another, generates possibly new disjuncts or discards disjuncts.

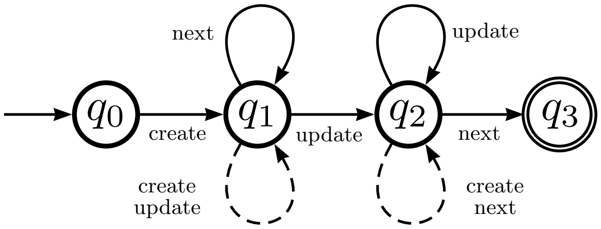

Figure 2: Finite automaton for safe enumeration tracematch of Figure 1

Figure 2 shows the automaton for the safe enumeration tracematch. As one can see, it looks very much like the most intuitive automaton for this pattern but augmented with additional loops (here dashed) on each non-initial and non-final state. Those loops here appear dashed, because they are of a special kind and have different semantics from usual edges. They are called *skip loops*.

The purpose of skip loops is to discard partial matches. The safe enumeration pattern is unfortunately one of the few where their relevance is somewhat hidden. Hence, in order to explain the purpose of skip loops, consider Figure 3. This figure shows the automaton for the tracematch *HasNext* which uses a pattern "next next" over a symbol alphabet {next,hasNext}.

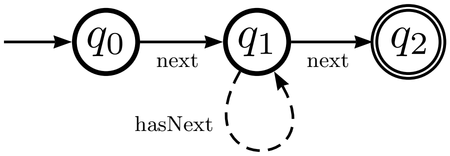

Figure 3: Finite automaton for tracematch pattern *HasNext*

The intent of this tracematch is to find all cases where there are two calls to Iterator .next(), with no call to hasNext() in between. Since the tracematch alphabet contains both the next and hasNext symbols, matching on the pattern "next next" implies that there was no call to hasNext() between the two next events.

This implicit negation is formulated in tracematches by including a symbol in the alphabet but not in the pattern, just like it is done with hasNext here. During runtime, when next() is called on a particular iterator i_1, a disjunct $\{i \mapsto i_1\}$ is generated on state q_1. Now, if another call to hasNext() follows, this binding can be discarded, because at least for the moment for this particular iterator i_1 the requirement is fulfilled. This is exactly what the skip loop on state q_1 achieves. When hasNext() is called on i_1, it *discards* the partial match for i_1 by deleting its disjunct from q_1. (An alternative implementation could move disjuncts back to the initial state, but discarding the disjunct saves memory.)

Running example To get a better feeling for the semantics of tracematches and the implications of our optimization, let us look at the following running example. Assume that we want to evaluate the safe enumeration tracematch over the code shown in Figure 4. The code does not do anything meaningful but it allows us to explain how tracematches work and which cases the different stages of our analysis can handle. In lines 5-10, the program modifies and iterates over the vector, vector, and does so in a safe way. In lines 12-15 it modifies and iterates over another vector, globalVec. It also calls doEvil (..) , modifying globalVec while the enumeration is used. This is a case which the tracematch should capture. In lines 17-18 a third vector and an enumeration over this vector are created.

The comments on the right-hand side of the figure label *allocation sites*, i.e. places where vectors or enumerations are allocated. We use those labels to denote objects. An object is labelled with the site at which it was allocated.

```
1  class Main {
2    Vector globalVector = new Vector();                                    //v2
3
4    void someMethod() {
5      Vector vector = new Vector();                                        //v1
6      vector.add("something");
7      for (Enumeration iter = vector.elements(); iter.hasMoreElements();) { //e1
8        Object o = iter.nextElement();
9        doSomething(o);
10     }
11
12     globalVector.add("something_else");
13     Enumeration it2 = globalVector.elements();                           //e2
14     doEvil(o);
15     it2.nextElement();
16
17     Vector copyVector = new Vector(globalVec);                           //v3
18     Enumeration it3 = copyVector.elements();                             //e3
19   }
20
21   void doSomething(Object o)
22   { /* does not touch globalVector */ }
23
24   void doEvil(Object o)
25   { globalVector.remove(o); }
26 }
```

Figure 4: An example program

In our static analysis, we attempt to remove unnecessary instrumentation points in the base program that trigger the tracematch at a point where it can statically be decided that the particular event can never be part of a complete match. Such instrumentation points are commonly called *shadows* in aspect-oriented programming [13, 21], and hence we will also use that term in the remainder of this paper. To see how one could identify unnecessary shadows, let us first manually find such places in the code for our running example.

Shadows occur wherever a tracematch symbol matches a part of the program. In our example, this means that we have shadows at each creation of an enumeration, each update of a vector and each call to Enumeration.nextElement(). However, when looking at the code more carefully, it should become clear that not all of the shadows are necessary for the example program.

In particular, the first sequence of statements in the lines 5-10 is safe in the sense that the pair of vector and enumeration is used correctly and the tracematch will not be triggered. Consequently, no shadows need to be inserted for this part of the program. Lines 12 to 15 and line 25 show an unsafe enumeration that should trigger the tracematch. So generally, shadows here need to stay in place. However, looking at the code more carefully, one can see that actually the shadow at line 12 is also superfluous, because the match that triggers the tracematch does not start before line 13, where the enumeration is actually created. In lines 17 to 18 we have a pair of vector and enumeration where the vector is never even updated. For this piece of code it should be obvious that no shadows are required.

In the next section we describe our static program analyses which automatically identify the unnecessary shadows.

3 Staged analysis

Our analysis is implemented using the *reweaving* framework [4] in abc. The basic idea is that the compiler first determines all shadows, i.e. all points in the program where instrumentation should be woven. This procedure returns what we call a *weaving plan*. This plan tells the weaver what needs to be woven at which shadows. In order to determine which shadows are unnecessary, a first weaving is done according to the original weaving plan. This results in a woven program on which our proposed staged analyses are performed. The analysis determines which shadows are unnecessary and removes them from the weaving plan. The program is then rewoven according to this new plan, resulting in a more efficient woven program.

The analyses are performed on the Jimple[1] representation of the woven program. In this representation, all instructions corresponding to tracematch shadows are specially marked so that they can be recognized.

An outline of the staged analyses is shown in Figure 5. Each stage uses its own abstract representation of the program and applies an analysis to this representation in order to find unnecessary shadows. After each stage, those shadows are removed so that subsequent stages do not have to consider them any more in their analyses.

The crucial point of this approach is that the earlier stages (on the top of the figure) are more coarse-grained than later ones. Hence they use a more lightweight abstract representation of the program and execute much faster. By applying stages in this order we make sure that at each stage only those shadows remain active which could not be proven unnecessary using an easier approach.

Figure 5 shows the three analysis stages we apply here as boxes. First we apply a quick check that determines if a tracematch can apply to a given program at all, just by looking at shadow counters, which are already computed during the initial weaving process. The second stage uses points-to information in order to find groups of shadows which could during runtime possibly lead to a complete match by possibly referring to a consistent variable binding. The third and final stage is flow-sensitive, meaning that we look at all those groups of shadows and try to determine in which order their shadows could possibly be executed when the program is run. In many cases, all shadows might already be removed in an early stage. When this happens, later stages are not executed at all. In any case, however, the code is eventually rewoven using the updated weaving plan, i.e. weaving only those shadows that have not been disabled before.

As the figure suggests, in general it can make sense to iterate the analysis and reweaving phases. In our experiments for this paper, we used however empty bodies for all tracematches, simply because we were only interested in the cost of matching, not executing a tracematch. If all tracematch bodies are empty, the tracematches themselves can trigger no joinpoints and hence their removal does not affect the analysis result

[1] Jimple is a fully typed three-address code representation of Java bytecode provided by Soot [28], which is an integral part of abc.

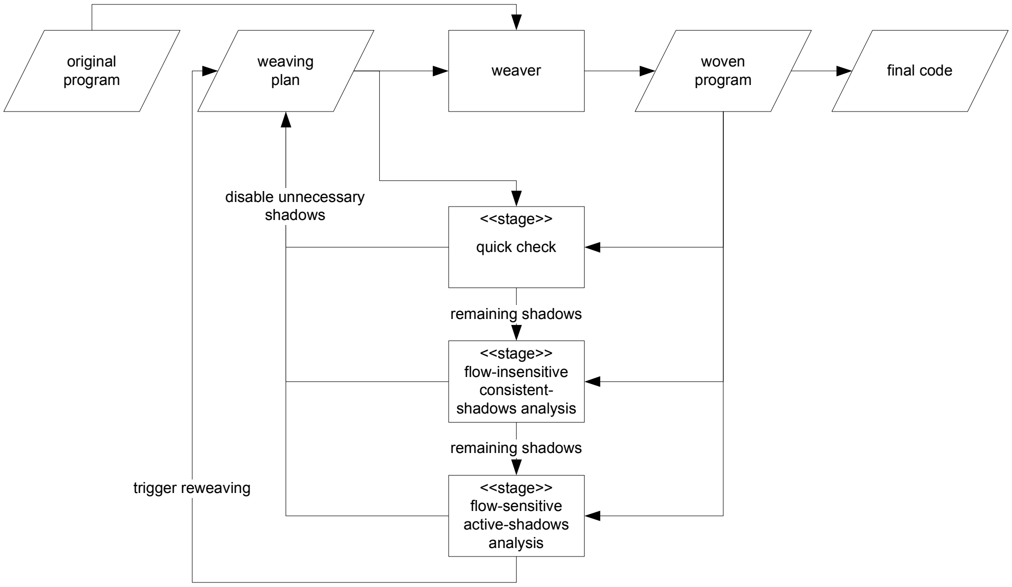

Figure 5: Outline of the staged analysis

in the first iteration.

In the following subsections, we explain all three stages as well as their required program abstractions in more detail.

3.1 Quick check

One use of tracematches is to specify behavioural constraints for Java interfaces. When developing a library, for example, one could ship it together with a set of tracematches in order to enforce that objects of that library are used in a certain way or in certain combinations. Consequently, it might often be the case that certain tracematches might never match or that only some of their symbols match, simply because the client uses only parts of the library.

For example, imagine a program which uses vectors but no enumerations. In this case, when applying the safe enumeration tracematch, `abc` would normally instrument all locations where a vector is updated, although an analysis of the whole program would show that the tracematch can never match.

The abstract program representation used by the quick check is simply a mapping from tracematch symbols to number of shadows at which the tracematch symbol may match. Those numbers are obtained during the initial weaving phase. For our running example, we would obtain the following mapping because enumerations are created at three places, they are advanced at two places and vectors are updated at three places.

$$\{create \mapsto 3, next \mapsto 2, update \mapsto 3\}$$

We use these counts, plus the tracematch automaton to determine if the tracematch could ever match. The key idea is that if a symbol that is necessary to reach a final state in the automaton has a count of 0 (i.e. no instances in the program under analysis exist), then there is no possibility that the tracematch could match.

We implement this check as follows. For each tracematch, we remove edges from its automaton whose label has a shadow count of 0. Then we check to see if a final state can still be reached. If the final state can't be reached, the entire tracematch is removed and all its associated shadows are disabled.

If the quick check fails for a tracematch, i.e. all necessary symbols were applied at least once, we have to change to a more detailed level of abstraction which leads us to the flow-insensitive analysis.

3.2 Flow-insensitive consistent-shadows analysis

A tracematch can only match a trace if the trace refers to symbols with *consistent variable bindings.* In the quick check we just used the names of the symbols and did not use any information about variable bindings. In contrast, the flow-insensitive *consistent-shadows analysis* uses points-to analysis results to determine when shadows cannot refer to the same object and thus cannot lead to consistent variable bindings. The analysis is flow-insensitive in the sense that we do not consider the order in which the shadows execute.

3.2.1 Preparation:

In order to prepare for this analysis, we first need points-to information for each variable involved in the tracematches. We compute the required points-to information as follows.

First we build a call graph using the Soot/abc internal Spark [18] framework. Spark builds a call graph for the whole program on-the-fly, i.e. by computing points-to information at the same time as discovering new call edges due to new points-to relationships. This first phase results in a complete call graph and context-insensitive points-to information for the whole program.

In our preliminary experiments we found that the context-insensitive points-to analysis was not always precise enough, and so we added a second phase that computes context-sensitive points-to analysis for those variables bound by shadows. For this second phase we use Sridharan and Bodík's demand-driven refinement analysis for points-to sets [23]. This algorithm starts with the call graph and context-insensitive results from the first phase and computes context information for *a given set of variables*, often yielding more precise points-to information for these variables. The advantage of this approach is that we need to perform this rather expensive computation only for variables that are really bound by shadows. In all our benchmarks this was fewer than 5% of the total number of variables. (For exact numbers, see Section 4.)

Our running example illustrates quite clearly why context-sensitive points-to analysis is required. In this case, context information is necessary to distinguish the different enumerations from each other. Since all are created within the factory method elements(), without such context-sensitivity, all enumerations would be modelled as the same abstract object — their common creation site inside the method elements(). Allocation sites e1, e2 and e3 would collapse, and so the analysis would have to assume that all three enumerations might actually be one and the same, penalizing the opportunities for shadow removal.

3.2.2 Building path infos:

At runtime, a tracematch matches when a sequence of events is executed which is matched by the given regular expression, however *only* if those events occurred with a *consistent variable binding.* The idea of the flow-insensitive analysis stage is to identify groups of shadows which could potentially lead to such a consistent variable binding at runtime.

At runtime, a final state in the tracematch automaton can be reached from any initial state, generally over multiple paths. A first observation is that edges which originate from symbols within a Kleene-* sub-expression are always optional. For example, in the safe enumeration tracematch (Figure 1), the initial next∗ *may*, but does not have to, match a joinpoint in order for a sequence to lead to a complete match. Hence, we first generate an automaton using a customized Thompson construction [27] that omits "starred" sub-expressions, modelling them with an ϵ-edges (which are then later on inlined).

Figure 6 shows the fail safe enumeration automaton after this transformation. We call this representation the *reduced* automaton. Note that skip loops are preserved in this representation, however no other strongly-connected components remain. Hence, we can enumerate all paths through this automaton which do not lead through a skip loop.

Then, for each such path we compute a *path info.* A path info consists of two components. The first holds information about which symbols the edges on the path are labelled with. The second records all labels of skip-loops that are attached to states on that path. For the labels of non-skip edges, we will later on also need the information of how often such a label occurs on the path. This yields the following definition.

Definition 1 (Path info)**.** Let *path* be a path from an initial to a final state in the reduced automaton.

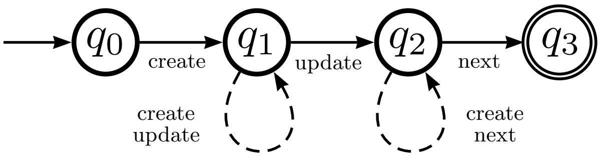

Figure 6: Automaton from Figure 2 with loops due to Kleene-* sub-expressions removed

A path info *info* consists of a set *skip-labels(info)* and a multi-set *labels(info)*, defined as follows. Assume we define for each state q the set *skip-labels*(q) as the set of labels of all skip-loops on q. Then, if $path = (p_1, l_1, q_1) \dots (p_n, l_n, q_n)$, then

$$labels := \biguplus_{1 \leq i \leq n} \{l_i\}$$

$$skip\text{-}labels := \bigcup_{1 \leq i \leq n} (skip\text{-}labels(p_i) \cup skip\text{-}labels(q_i))$$

where $\biguplus$ denotes the union for multi-sets. (A multi-set or bag is a similar to a set but can hold the same object multiple times.) In the following, we denote multi-sets with square brackets of the form $[a, a, b]$.

For a tracematch *tm*, we denote the set of all its path infos by *infos(tm)*. It is defined as the set of all path infos for all paths through its reduced automaton.

For the fail safe enumeration tracematch in our example, only one path exists: $(q0, \text{create}, q1), (q1, \text{update}, q2), (q2, \text{next}, q3)$. Hence the set *infos* has the following form.

$$\begin{aligned} &infos(\text{FailSafeEnum}) = \\ &\quad \{(\ labels = [\text{create}, \text{update}, \text{next}],\ skip\text{-}labels = \{\text{create}, \text{update}, \text{next}\}\ )\} \end{aligned}$$

The reader should not be misled by this example. In general, *labels* and *skip-labels* do not have to coincide. For example, for the automaton in Figure 3, we would have a single path info with $labels = [\text{next}, \text{next}]$ and $skip\text{-}labels = \{\text{hasNext}\}$.

3.2.3 Building groups of shadows with possibly consistent binding:

With the path infos computed, we have information about what combinations of shadows are *required* for a complete match. In the next step we try to find groups of shadows that fulfil this requirement. This means that we look for groups of shadows which contain the labels of the *labels* field of a path info and, in addition, share a possibly consistent binding. But before we define shadow groups, let us first formally define how a single shadow is modelled.

Definition 2 (Shadow)**.** A shadow s of a tracematch *tm* is a pair $(lab_s, bind_s)$ where lab_s is the label of a declared symbol of *tm* and $bind_s$ is a variable binding, modelled as a mapping from variables to points-to sets. In the following we assume that the mapping $bind_s$ is extended to a total function that maps each variable to the full points-to set $\top$ if no other binding is defined:

$$bind_s(v) := \begin{cases} bind_s(v), \text{if } bind_s(v) \text{ explicitly defined} \\ \top, \text{otherwise} \end{cases}$$

Here, $\top$ is defined as the points-to set for which holds that for all points-to sets $s : s \cap \top = s$.

In our running example, the update shadow in line 6 would be denoted by $(\text{update}, \{v \mapsto \{v1\}\})$ as the only objects v can point to are objects being created at creation site $v1$.

Definition 3 (Shadow group)**.** A *shadow group* is a pair of a multi-set of shadows called *label-shadows* and a set of shadows called *skip-shadows*. We call a shadow group *complete* if it holds that: (1) its set of labels of *label-shadows* contains all labels of a path info of a given tracematch; and (2) its set of *skip-shadows* contains all shadows which have the label of a skip loop of a state on this path and a points-to set that overlaps with the one of a label shadow.

This definition implies that a complete shadow group has: (1) enough shadows in its *label-shadows* to drive a tracematch state machine into a final state; and (2) that all shadows that could interfere with a match via skip loops are contained in *skip-shadows*.

Definition 4 (Consistent shadow group)**.** A *consistent* shadow group g is a shadow group for which all variable bindings of all shadows in the group have overlapping points-to sets for each variable. More formally, if $vars$ is the set of all variables of all shadows in g, then it must hold that:

$$\forall s_1, s_2 \in (\textit{label-shadows} \cup \textit{skip-shadows})\ \forall v \in vars:\ bind_{s_1}(v) \cap bind_{s_2}(v) \neq \emptyset$$

Conceptually, a complete and consistent shadow group is the static representation of a possibly complete match at runtime. For such a shadow group, there is a possibility that if the label shadows in this group are executed in a particular order at runtime, the related tracematch could match. Skip shadows in the same group could prevent such a match when executed.

In particular, if a shadow group has a multi-set of label shadows which is *not* consistent this means that no matter in which order those shadows are executed at runtime, this group of shadows can *never* lead to a complete match. Consequently, we can safely disable all shadows which are not part of any consistent shadow group. The algorithm we use in order to compute all complete and consistent shadow groups is shown as Algorithm 1. (While this algorithm first constructs all shadow groups and then filters out inconsistent ones, our actual implementation is a little smarter, taking care that inconsistent groups are even not at all computed in the first place.)

Based on the consistent shadow groups, flow-insensitive shadow removal is then quite easy. For each shadow that exists in the program, we look up if it is member of at least one consistent shadow group (i.e. it is either a label-shadow or a skip shadow of that group). If this is *not* the case, the shadow can never be part of a complete, consistent match and can safely be removed.

In our running example, this is true for the shadow in line 18. Since for this create-shadow there exists neither an update shadow for the same vector nor a next-shadow for the same enumeration, there can no complete and consistent shadow set be computed that contains the create-shadow.

Here we can also see that context information for points-to sets is important. As noted earlier, without context information, all enumerations would be modelled by the same abstract object. Hence, in this case, the points-to sets for those shadows would overlap and the shadow in line 18 *could* be part of a complete and consistent match, in combination with one of the vectors globalVector or vector.

If after this stage there are still shadows remaining we know that there exist groups of shadows which have a possibly consistent variable binding. This means that if such shadows are executed in a particular order at runtime, the related tracematch could indeed be triggered. Hence, it is only natural that in the next stage we compute information that tells us whether those shadows could can actually be executed in the required order or not. This leads us to the flow-sensitive consistent-shadows analysis stage.

3.3 Flow-sensitive active-shadows analysis

As input to this stage we expect a set of complete and consistent shadow groups as well as a complete call graph, both of which were already computed earlier. (In the following, when we refer to a shadow group, we always assume it is complete and consistent.)

In order to determine in which order shadows could be executed during runtime, we need a flow-sensitive representation of the entire program. It is a challenge to build such a representation efficiently. Since any Java program is potentially multi-threaded, we also have to take into account that shadows could be executed by multiple threads. This makes it more difficult to determine whether a shadow may run before or after another.

A tracematch can be defined to be *per-thread* or *global.* For a per-thread tracematch, a separate automaton is executed for each thread, and only events from that one thread affect the automaton. A global tracematch is implemented using a single automaton which processes events from all threads. Hence, for global tracematches, our analysis must handle multi-threading soundly.

Algorithm 1 complete and consistent shadow groups

```
$shadowgroups := \emptyset$
**for** tracematch $tm$ **do**
  **for** each path info $info$ in $infos(tm)$ **do**
    /* cross product over all shadows for all labels in the path info */
    $crossProduct := \bigotimes_{l \in labels(info)} [s \in shadows(tm) \mid label(s) = l]$
    /* filter out results with non-overlapping variable binding */
    **for** $bag \in crossProduct$ **do**
      **for** $s_1, s_2 \in bag$ **do**
        **for** $var \in \{v \mid v$ bound by $bind_{s_1} \vee v$ bound by $bind_{s_2}\}$ **do**
          **if** $bind_{s_1}(var) \cap bind_{s_2}(var) = \emptyset$ **then**
            $crossProduct := crossProduct - \{bag\}$
          **end if**
        **end for**
      **end for**
    **end for**
    /* find skip shadows */
    **for** $bag \in crossProduct$ **do**
      $skipshadows := \emptyset$
      **for** label $l$ in $skip\text{-}labels(info)$ **do**
        **for** skip shadow $s_s \in \{s \in shadows(tm) \mid label(s) = l\}$ **do**
          **for** label shadow $s_l \in bag$ **do**
            $overlaps := true$
            **for** variable $v$ bound by $s_s$ **do**
              **if** $bind(s_s) \cap bind(s_l) = \emptyset$ **then**
                $overlaps := false$
              **end if**
            **end for**
            **if** $overlaps$ **then**
              $skipshadows := skipshadows \cup \{s_s\}$
            **end if**
          **end for**
        **end for**
      **end for**
      /* add new shadow group as pair of label and skip shadows */
      $shadowgroups = shadowgroups \cup \{(bag, skipshadows)\}$
    **end for**
  **end for**
**end for**
```

Also, a whole program abstraction may potentially be very large. There might potentially be thousands of shadows spread over hundreds of methods. Hence it is important that we keep our program abstraction concise at all times.

3.3.1 Handling of multi-threading

We handle the first problem of multi-threading conservatively. In the preparation phase for the flow-insensitive analysis stage, we already constructed a complete call graph. In this call graph, call edges that spawn a thread are already specially marked. Using this information, we can easily determine by which threads a given shadow can be executed.

Then, in an initial preprocessing step, we filter the list of all shadow groups in the following way. If a shadow group is associated with a global tracematch and contains shadows which are possibly executed by multiple threads, we "lock" all its shadows (i.e. they will never be removed, not by this stage nor by subsequent stages) and remove the group from the list. The locking makes the analysis conservative with respect to threads. For the resulting list of shadow groups we then know that all shadows contained in a group are only executed by a single thread each. Hence, no additional treatment of multi-threading is necessary.

3.3.2 A flow-sensitive whole-program representation

In the next step, we build a flow-sensitive representation of the whole program. Such a representation naturally has to depend on the static call graph of the program.

Call graph filtering In order to adhere to our principle of keeping our abstraction as small as possible at all times, we first filter this call graph in the following way. If in the call graph there is an outgoing call edge in whose transitive closure there is never any method of interest reachable (i.e. a method that contains a shadow), this edge and its entire transitive closure is removed.

Per-method state machines For each method that remains in this filtered call graph, we know that either it is "interesting" because it contains a shadow or it calls another interesting method. For those methods we do need control flow information, i.e. information about the order in which shadows may be executed during runtime and in which other methods may be called.

We encode such flow-information by a finite state machine for each such method. In order to generate this abstraction for a particular method, we first generate a control-flow graph for the method. This graph encodes the entire possible control flow, including exceptional one. Each node in this graph represents a statement.

We generate the per-method state machine by turning edges in the control-flow graph into transitions in the state machine. For each statement that is neither an invoke statement nor contains a shadow, we generate an ϵ transition. For each statement that contains a shadow s, we generate an edge labelled with s. For each statement that contains an invoke expression ie, we generate an *invoke-edge* labelled with all possible call targets of ie. Since our call graph is filtered, there may be invoke expressions for which there is no target. In this case, we label the edge with ϵ. If an entire method *body* contains a shadow s (e.g. an **execution** shadow), we generate an edge labelled with s which, in the case of a *before* symbol, precedes all other edges in this state machine or, in the case of an *after* symbol follows all other edges.

Before we proceed, we then inline all epsilon transitions in those state machines and minimize them one by one by using standard algorithms known from automata theory. Then, in a last step, we equip each state machine with a unique start and end state. The start state is connected to all initial states via ϵ-transitions and so is any final state to the end state. (Since we do not use the state machine for language acceptance, the notion of a final state is for our purposes not important in any other way.)

Inter-procedural combination In order to obtain an abstract representation of the entire program, we then combine all the per-method state machines inter-procedurally. For each invoke-edge we generate ϵ transitions from the source state of this edge to each start state of each possible call target. We generate a second transition from each end state of those call targets to the target state of the original invoke-edge. Finally, the original invoke-edge is removed.

Figure 7 shows this step for our running example. This figure shows three state machines for the methods doSomething(), someMethod() and doEvil() and how they become inter-procedurally combined using ϵ-transitions.

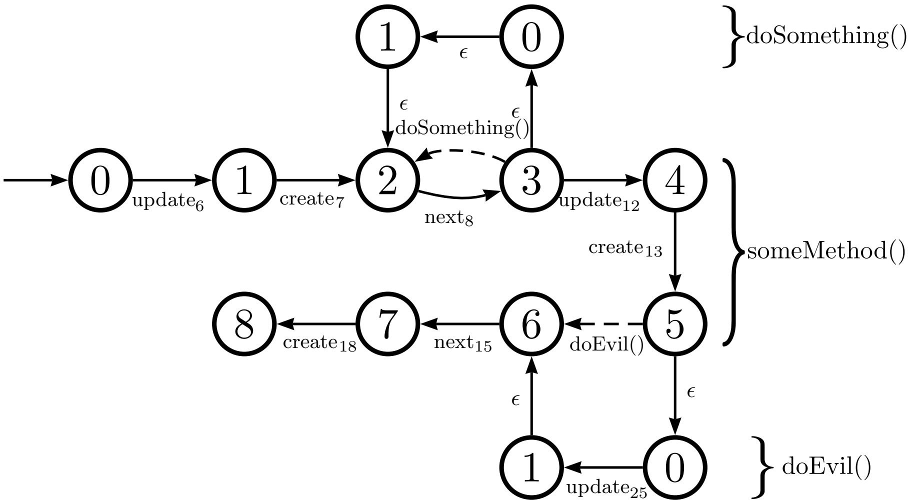

Figure 7: complete state machine for the running example during construction; for illustrative purposes, the shadow labels are attached with their respective line numbers

Note that this way of combining automata is context-insensitive. In the resulting automaton there exist more paths than are actually realizable at runtime. One could branch out from a call statement c_1 to a possible call target t but then return to another caller c_2 of the same call target. This way of automaton construction is relatively cheap but gives away precision, as Section 4 will show.

Abstract interpretation via fixed point iteration This whole-program state machine is the input to our actual flow-sensitive analysis. The task of this analysis is to compute if some part of this state machine contains such a path that when executing the program along this path at runtime, the tracematch could match. To us, it appeared that the most sensible way to do so is to perform a complete abstract interpretation of the actual tracematch machinery.

This abstract interpretation evaluates an *abstract counterpart* of the actual tracematch automaton (i.e. the one that is evaluated at runtime) over the whole-program state machine. Since the latter can have cycles, we employ, as is usually done in data-flow analysis [15], a fixed-point iteration.

The only two differences of the abstract interpretation in comparison to the evaluation at runtime are the following. Firstly, wherever the actual implementation binds variables to objects, the abstract interpretation binds them to points-to sets. Consequently, where at runtime, the implementation checks for reference equality, the abstract interpretation checks for overlapping points-to sets. In the case of skip loops, variable bindings are not updated at all (due to the lack of must-alias information, see below).

The other difference is that while during runtime, the implementation exposes no explicit information about where partial matches occurred, the static abstraction needs to determine which shadows were visited on the way to a final state. Hence, in each disjunct, we store an additional *history* component: the set of shadows which this disjunct was propagated through. When such a disjunct reaches a final state, we can inspect its history and so determine which shadows need to be kept active in order to trigger the match for this disjunct at runtime. The history is also updated in case a skip loop is visited.

We start off with an initial tracematch configuration in the unique initial state of this whole-program state machine, which represents the fact that when the program starts, the tracematch is in its initial configuration. In terms of Figure 7, this would associate the following configuration with the initial node of the whole-program state machine.

$$(q_0 \mapsto \textbf{true}, q_1 \mapsto \textbf{false}, q_2 \mapsto \textbf{false}, q_3 \mapsto \textbf{false})\}$$

Here **true** is the constraint $\{(\emptyset, \emptyset)\}$ consisting of a single disjunct with empty variable binding and history while **false** is the empty constraint (modelled by the empty set of disjuncts).

This configuration is then driven through the whole-program state machine until a fixed-point is reached. Whenever a disjunct is propagated, its history component is updated with the shadow that triggered the propagation. Due to internals of the tracematch machinery, this is only the case if a constraint really moves to a new state. For example at node number 1 (of method someMethod()) in Figure 7, the configuration is still same the initial configuration as above. At node number 2, one would get

$$(q_0 \mapsto \textbf{true}, q_1 \mapsto \{(\{v \mapsto v_1, e \mapsto e_1\}, \{create_7\})\}, q_2 \mapsto \textbf{false}, q_3 \mapsto \textbf{false})\}$$

stating that the abstract tracematch automaton has one single partial match in state 1 with a variable mapping of $\{v \mapsto v_1, e \mapsto e_1\}$ which was produced by shadow $create_7$.

At merge-points (here only the same node number 2), configurations are merged by joining their constraints per state, i.e. two constraints with mappings $q_i \mapsto \{d_1, d_2\}$ and $q_i \mapsto \{d_2, d_3\}$ (for disjuncts d_1, d_2, d_3) is merged to a constraint with mapping $q_i \mapsto \{d_1, d_2, d_3\}$.

During the computation of the fixed point, whenever a disjunct reaches a final state of a configuration, we copy its history to a *global set* of *active shadows*. When the fixed point is reached, we know that all shadows in this set may lead to a complete match, with the binding that is stored in the disjunct, and hence have to be retained. All shadows which are never added to this set during the fixed point computation can safely be discarded.

Performance improvements The aforementioned fixed point computation generally works but it might not be very efficient. Hence, we apply two different performance optimizations, one of which does not sacrifice precision and one of which does.

The general problem is that the set of possible different disjuncts is quite large. Consequently it might take a long time and a lot of memory before the fixed point is reached. In the example, the whole-program state machine has only 8 non-ϵ edges. In realistic benchmarks, this number can however well be around 1000.

If this happens, two different cases can occur. In the first case (the good case), points-to sets are relatively precise. Then, we can apply the following trick to "shrink" the whole-program state machine without losing precision. We simply do not perform the fixed point iteration once but rather once for each shadow group. In each such iteration, all edges with shadows which are not part of the group are treated as ϵ transitions, simply copying configurations without modification. (This leads back to the idea of the declarative semantics of tracematches where multiple automata are executed simultaneously, one for each binding.)

In the running example, this would mean that first we evaluate only the sub-automaton on the left-hand side of node 3, then the one between nodes 3 and 7 and then the one between nodes 7 and 8. As a consequence, disjuncts "travel" smaller distances through the graph and so we gain disjuncts with smaller history components, hence as a consequence less disjuncts in general.

In the second case (the hard one), points-to sets overlap a lot, either because indeed the same objects trigger shadows over and over again or because the computed points-to sets are imprecise. This means that each shadow group is relatively large. In particular this is the case for the *skip-shadows* component of the shadow group; the *label-shadows* component has bounded size by construction. Hence, the above trick does not help because still in each iteration a lot of edges would be enabled.

In this case, we apply the same trick as for multi-threading. We lock all shadows in the *skip-shadows* set, i.e. they will not be removed from the weaving plan. Because this is the case, we can then disable edges

pattern name	description
ASyncIteration	only iterate a synchronized collection c when owning a lock on c
FailSafeEnum	do not update a vector while iterating over it
FailSafeIter	do not update a collection while iterating over it
HashMap	do not change an object's hash code while it is in a hash map
HasNextElem	always call hasNextElem before calling nextElement on an Enumeration
HasNext	always call hasNext before calling next on an Iterator
LeakingSync	only access a synchronized collection using its synchronized wrapper
Reader	don't use a Reader after it's InputStream was closed
Writer	don't use a Writer after it's OutputStream was closed

Table I: description of tracematch patterns

labelled with those shadows during fixed point iteration. In our benchmarks, we switched to this evaluation mode if $|\textit{skip-shadows}| > 2 * |\textit{label-shadows}|$.

The combination of those two techniques allowed us to compute the fixed point in all our benchmarks within a few minutes. However, it often gives away crucial precision, as we will see in Section 4.

Handling of skip loops One important issue that has not yet been explained is the handling of skip loops. As explained earlier, the purpose of a skip-loop is to discard partial matches under certain circumstances. In the example we gave in Section 2, this is the case when a disjunct of the form $\{i \mapsto i_1\}$ exists and then hasNext is invoked on the iterator i_1.

At runtime, we can remove this partial match because we know that for the object i_1 the property is currently fulfilled. The automaton can be "reset" to its initial configuration with respect to i_1. At compile time, we are only allowed to apply the same strong update of the automaton constraints if we know for sure that the object referenced by variable i at the hasNext event *must* be the same as the object referenced by i as the previous next event. In other words, we have to know if the references to variable i at both locations in the code *must* be aliased.

As there is currently no must-alias analysis in Soot, we perform a weak update on skip-loops that does not discard partial matches. Unfortunately, this makes it impossible to rule out patterns as the one mentioned above merely by looking at the possible execution order. (Still, we can rule out skip-loops that are only executed on paths that never lead to a final state.) Our next phase of work will investigate the kinds of must-alias analyses we need to handle skip loops more precisely. Fink et al. show in their work [11] what a general solution could look like for the case of typestates, where one only reasons about one object at a time.

4 Benchmarks

In order to evaluate the feasibility and effectiveness of our approach we applied our analysis to a combination of nine different tracematches, applied to version 2006-10 of the DaCapo benchmark suite [7]. The tracematches validate generic safety and liveness properties over common data structures in the Java runtime library. They are briefly described in Table I. As usual, all our benchmarks are available on `http://www.aspectbench.org/`, along with a version of `abc` implementing our optimization. In the near future we also plan to integrate the analysis into the main `abc` build stream.

The reader should note that we chose some of our tracematches because we knew they would be particularly challenging. For example, the *HashMap* tracematch binds a hash code, which is an integer value. We found it interesting to see what effect the presence of a non-pointer variable would have on our pointer-based analyses. The *ASyncIteration* benchmark uses an **if**-pointcut accessing the native method Thread.holdsLock(Object). This is challenging because there is no chance to generally evaluate such a pointcut at compile time. The question is whether this fact generally impedes the analysis or not.

benchmark	no tracematch	ASyncIteration	FailSafeEnum	FailSafeIter	HashMap	HasNextElem	HasNext	LeakingSync	Reader	Writer
antlr	4098	1.42	2.20	0.93	0.44	6.54	-0.15	**25.28**	**966.98**	**108.76**
bloat	9348	**99.17**	0.75	**>8h**	**139.08**	0.58	**3872.66**	**497.35**	-2.95	**92.52**
chart	13646	0.39	0.01	**20.73**	0.15	0.13	0.99	**345.30**	0.32	0.29
eclipse	50003	2.36	1.10	3.44	2.36	0.53	4.81	2.61	0.28	-1.21
fop	3102	-9.96	-8.67	1.06	5.35	-4.13	-9.93	**589.30**	0.71	6.74
hsqldb	12322	0.00	-0.32	0.03	0.19	0.07	-0.16	0.79	0.32	-0.06
jython	11133	1.47	2.04	6.57	1.05	-1.17	2.67	-11.17	-0.89	0.50
lucene	17068	1.29	**30.36**	9.57	3.40	**17.17**	2.22	**422.52**	1.78	1.12
pmd	12977	2.96	-0.11	**157.61**	-0.83	-1.85	**158.23**	**31.26**	2.43	-0.21
xalan	13083	1.86	0.20	0.71	1.35	-0.41	2.34	4.20	1.70	0.47

Table II: Runtime overheads of the benchmarks before applying our optimizations

The tracematches *HasNext* and *HasNextElem* specify properties where something *must always* happen in order to avoid a match. After a call to next(), hasNext() must be called before the next call to next(). (In the verification community, such properties are often called *liveness properties* [16].) As mentioned in Section 3.3.2, the flow-sensitive analysis cannot remove shadows for such properties without using must-alias information. The flow-insensitive analysis would also perform badly on those particular properties, simply because on virtually every iterator hasNext() is called if and only if next() is called. Hence, for those benchmarks we expected a very low shadow removal ratio. Yet, the benchmarks helped us to validate the completeness of our implementation because we knew that in those cases neither the flow-insensitive nor the flow-sensitive stage should remove any shadows.

For our experiments we used the IBM J9 JVM version 1.5.0 SR3 (64bit) with 2GB RAM on a machine with AMD Athlon 64 X2 Dual Core Processor 3800+. We used the `-converge` option of the DaCapo suite which runs each benchmark multiple times to assure that the reported execution times are within a confidence interval of 3%.

Table II shows the run times of the benchmarks without our optimizations, but with the optimizations mentioned in [5] already enabled. The leftmost column shows the benchmark name, then follows the raw runtime with no tracematch present (in milliseconds). The other columns show the overheads in percent over this baseline. We marked all cells with values larger than 10% as boldface, values within the confidence interval appear gray. The benchmark *bloat/FailSafeIter* was stopped after around 8 hours of benchmarking time. This benchmark is very hard to handle, dynamically as well as statically, because it makes extraordinarily heavy use of long-lived iterators and collections. We shall return to this benchmark later, when we discuss the precision of our analysis.

As we can see from the table, some benchmarks expose a significant overhead.[2] In these cases the whole program optimizations presented in this paper are worth applying. In particular, given the sometimes large runtime overhead, the programmer might well want to trade some of this overhead for compile time.

We applied our analysis to all 90 benchmarks, and in Table III we report on the improvements for the 18 interesting cases with an overhead of more than 10%. We captured the optimized program after each stage in order to see how many shadows were removed and are still remaining and in order to evaluate the runtime impact of the shadow removal for that stage. The table shows per benchmark five different shadow counts: all shadows, shadows remaining after the quick check, reachable shadows remaining after call graph

[2] The speedups for fop and jython apparently originate from the fact that those benchmarks are *bistable*. Depending on scheduling order they settle down in one of two different highly predictable states. Additional instrumentation can sometimes affect this order and make the benchmark settle into a more favourable state, i.e. make the benchmark execute faster. This interpretation was suggested by Robin Garner, one of the developers of the DaCapo benchmark suite.

#	benchmark	all	quick	reachable	flow-ins.	flow-sens.	final stage
1	antlr/LeakingSync	170	0	0	0	0	quick check
2	antlr/Writer	56	0	0	0	0	quick check
3	bloat/ASyncIteration	419	0	0	0	0	quick check
4	bloat/LeakingSync	2145	0	0	0	0	quick check
5	chart/LeakingSync	920	0	0	0	0	quick check
6	fop/LeakingSync	2347	0	0	0	0	quick check
7	pmd/LeakingSync	986	0	0	0	0	quick check
8	lucene/LeakingSync	653	653	294	0	0	flow-ins.
9	antlr/Reader	53	53	46	15	15	flow-sens.
10	bloat/HashMap	57	57	16	2	2	flow-sens.
11	bloat/Writer	206	206	87	8	8	flow-sens.
12	lucene/FailSafeEnum	61	61	41	5	5	flow-sens.
13	pmd/FailSafeIter	529	529	129	90	90	flow-sens.
14	chart/FailSafeIter	469	469	105	105	105	flow-sens.
15	lucene/HasNextElem	22	22	14	14	14	flow-sens.
16	pmd/HasNext	346	346	87	86	86	flow-sens.
17	bloat/FailSafeIter	1500	1500	1015	1015	1015	aborted
18	bloat/HasNext	947	947	639	639	639	aborted

Table III: Number of active shadows after applying each stage

construction (note that removing unreachable shadows has no impact on the runtime), and finally shadows remaining after the last two analysis stages. The rightmost column shows the last stage that was applied.

The table is split vertically into multiple parts. For the benchmarks in the first part (rows 1-7), the quick check was able to eliminate all shadows. For row 8, the flow-insensitive analysis removed all 294 reachable shadows. In the benchmarks in rows 9-13, the flow-insensitive analysis removed at least some shadows, most often all but a few. In the benchmarks in row 14-16, the flow-insensitive analysis was ineffective. In benchmarks 17 and 18, the analysis failed to complete in a reasonable time or aborted due to insufficient memory.

The results show that the quick-check is very effective, removing all shadows in seven benchmarks. The flow-insensitive stage is generally very effective too, reducing the instrumentation and runtime overhead in another seven cases. We wish to point out that even in the case of *bloat/HashMap*, where primitive values are bound, the flow-insensitive analysis can still rule out many shadows by relating those remaining variables which bind objects. In one case (number 8), it is even able to prove the program correct, i.e. that all synchronized collections are only accessed via their synchronized wrapper.

The reader should note that optimizations as we propose here would be hopeless to perform on a a hand-coded monitor in plain AspectJ. Consequently at least in cases 1-8 where we remove all shadows, the optimized benchmark runs faster than it could ever be achieved using not tracematches but AspectJ only.

Looking at the flow-sensitive stage, we were very disappointed to see that it did not manage to remove more instrumentation over the flow-insensitive stage. While in some microbenchmarks which we used for testing, it yielded significant improvements, in the DaCapo benchmark suite it was not even able to remove a single additional shadow. We were able to identify three different factors that lead to this behaviour. We hope that these observations will lead to better analyses which can find further improvements.

Firstly, if a lot of shadows remain after the flow-insensitive analysis, this often indicates that for some reason there is a large overlap between points-to sets. When this is the case, it is however equally hard for the flow-sensitive analysis to tell different objects apart and hence to relate events on those objects temporally. As noted in Section 3.3, in such situations we often only perform a lightweight fixed point computation which treats skip shadows conservatively. In cases like *pmd/FailSafeIter* unfortunately, this seems to give away a lot of crucial precision.

#	benchmark	analysis	pre-opt.	quick	flow-ins.	flow-sens.
1	antlr/LeakingSync	< 0:01	25.28	0.15	-0.07	-1.00
2	antlr/Writer	< 0:01	108.76	3.44	4.00	2.76
3	bloat/ASyncIteration	< 0:01	99.17	18.44	18.68	18.59
4	bloat/LeakingSync	< 0:01	497.35	16.69	16.04	16.78
5	chart/LeakingSync	< 0:01	345.30	1.82	1.83	1.60
6	fop/LeakingSync	< 0:01	589.30	-9.16	-7.03	-9.77
7	pmd/LeakingSync	< 0:01	31.26	-0.73	-0.66	-1.09
8	lucene/LeakingSync	2:17	422.52	448.69	-4.04	-4.93
9	antlr/Reader	2:03	966.98	408.93	20.60	20.40
10	bloat/HashMap	7:02	139.08	134.11	2.57	3.61
11	bloat/Writer	7:34	92.52	280.03	4.11	3.59
12	lucene/FailSafeEnum	1:56	30.36	27.84	-1.80	-2.86
13	pmd/FailSafeIter	20:47	157.61	161.27	78.16	79.04
14	chart/FailSafeIter	7:52	20.73	20.52	22.36	20.56
15	lucene/HasNextElem	1:52	17.17	13.18	12.42	11.92
16	pmd/HasNext	4:20	158.23	167.73	169.08	158.13
17	bloat/FailSafeIter	aborted	307987.29	307987.29	n/a	n/a
18	bloat/HasNext	aborted	3872.66	3895.18	4013.53	n/a

Table IV: Runtimes of the analysis (left, in m:ss) and runtime overheads of the benchmarks (right, in percent) after applying each stage

Secondly, as we explained in Section 3.3.2, our whole-program state machine is context-insensitive, meaning that we over-approximate the set of actually realizable paths by not explicitly outgoing with returning call edges. This seems to lose precision in those cases where overlapping points-to sets are actually not the problem.

Thirdly, we handle multi-threading in a very conservative way. In benchmarks like lucene, the program does not trigger the tracematch only because it uses explicit wait/notify. Without analyzing such lock patterns explicitly, there is little hope for any more precision in those cases.

Case 14, *chart/FailSafeIter*, could also not be improved upon because of multi-threading. In addition, points-to sets largely overlapped due to the use of reflection which caused a safe over-approximation of points-to sets.

In the cases of *bloat/FailSafeIter* and *bloat/HasNext*, the analysis ran out of memory. The problem with bloat is[3] that it uses extraordinarily many iterator accesses and modifications of collections. In addition, almost all iterators and collections are very long-lived, so that points-to sets vastly overlap. The construction of the whole-program state machine suffers even more from the fact that bloat defines its own collections which delegate to collections of the Java runtime (JRE). Usually, collection classes are defined inside the JRE and thus not weavable and produce no shadows. Hence, due to the call graph abstraction, calls to hasNext() or updates to collections produce no edges in the whole-program state machine. In bloat, all those optimizations fail, making the problem of efficient model construction very hard to solve.

Table IV shows the runtimes of those 18 optimized benchmarks. As we can observe, there is most often a very direct relation between the number of shadows removed and the speedup gained by the optimization. After applying all three optimization stages, all but the benchmarks in rows 13 and 16-18 execute almost as fast as the un-instrumented program.

[3] Just before submitting the final version of this paper, through personal communication with Feng Chen [8] we found out that bloat within DaCapo apparently processes parts of itself as input (bloat is another bytecode transformation package). Hence, it might also be the case that our instrumentation modified bloat's input.

Per-thread tracematches We further analyzed per-thread versions of the tracematches *HasNext* and *HasNextElem* (in our eyes, the per-thread modifier makes no sense for the other configurations). Unfortunately, this seemed to yield no improvements in terms of precision and shortened the execution time of the analysis only marginally.

4.1 Execution time of the analysis

The analysis was run on the same machine configuration as the benchmarks but with a maximal heap space of 3GB. Total runtimes of the analysis are shown on the left hand side of Table IV. The longest successful analysis run we had was pmd/FailSafeIter with a total analysis time of almost 21 minutes. The different stages of this run are distributed as follows (m:ss).

- 0:01 - quick check
- 2:27 - call graph construction, points-to set creation, call graph abstraction
- 0:03 - flow-insensitive analysis
- 0:20 - creation of per-method state machines
- 1:48 - creation of whole-program state machine
- 15:51 - flow-sensitive fixed-point iteration

As we can see, the most expensive phase is the flow-sensitive fixed point iteration, followed by the time spent in the construction of the call graph and points-to sets. The quick-check is so fast that it is always worthwhile. The flow-insensitive analysis, in combination with its preparation phase, still runs in reasonable time. As Table IV shows, usually the runtime is between 3 and 10 times shorter.

It proved very sensible to make use of the demand-driven refinement-based points-to analysis. For example, in pmd/FailSafeIter, we queried points-to sets for 691 variables only, where a full context-sensitive points-to analysis would have had to compute context-sensitive points-to sets for all 33993 locals in the pmd benchmark.

The flow-sensitive analysis generally adds a large computational burden and our results show that it does not find any improvements over the cheaper flow-insensitive stage. We plan to further refine that phase in future work to see if it is really worthwhile pursuing.

5 Related work

While a lot of related work has been done in static program analysis and verification (model checking), there has been little previous work on using those techniques to speed up runtime monitoring. We list notable exceptions here.

Typestate properties Typestate properties [26] have been widely studied in the past. Very recently, Fink et al. presented a static optimization for runtime checking of typestate properties [11]. The analysis they present is in flavour similar to ours. In particular it is also implemented in multiple stages, one of which is flow-sensitive. However typestate properties allow one to express temporal constraints about *one single* object only, making their flow-insensitive and flow-sensitive stages simpler than ours. The authors paid special attention to the handling of strong updates, using must-alias information, which we leave to future work. The analysis Fink et al. present did not address the issue of multi-threading.

In terms of expressiveness, we believe that tracematches are equivalent to *generalized* typestate properties [17]. While normal typestate properties allow to reason about a single object each, generalized typestate properties allow to reason about multiple objects in combination. The only difference to tracematches seems to be the syntax (state machines vs. regular expressions).

PQL. In [20], Martin et al. present their Program Query Language, PQL. They experimented with a flow-insensitive analysis similar to our consistent-shadows analysis. However, their analysis is still not integrated within the PQL tool, making effective comparisons impossible at the current time. We suspect, that their flow-insensitive analysis performs very similarly to ours since they made similar design decisions. In particular they also do not take must-alias information into account. However, our analysis should in the general case be much faster, because unlike the analysis for PQL ours is staged, employing a very effective quick check first. Also we compute context for points-to sets for certain variables only while they apply this very expensive computation for all program variables.

History based pointcut languages Various other pointcut languages have been proposed that allow to match on histories of events, both in the aspect-oriented programming [9, 22, 29] and runtime verification community [8, 12, 24]. While we believe that for most such languages, depending on their expressiveness, similar analysis could be constructed, one crucial ingredient to the success of such a project is the use of an integrated compiler. For instance, one needs to be able to disable shadows that were proven unnecessary. Without access to an aspect-oriented compiler like `abc`, this seems almost impossible. Consequently we are not aware of any related work by other research groups on that topic, apart from the ones mentioned above.

6 Discussion and future work

In this work we have proposed a staged static analysis for reducing the overhead of finite-state monitors. We have presented three stages including a very coarse-grain and inexpensive quick check based only on shadows matching symbol names, a flow-insensitive consistent-shadows analysis that finds all shadows with consistent points-to sets, and a flow-sensitive active-shadows analysis that also takes into consideration the order in which shadows execute.

As is often the case in program analysis, we were somewhat surprised that the first two simpler stages were the most effective. The quick check analysis is very simple and also quite effective in eliminating tracematches that can never match a base program. We believe that this test will be very useful in situations where whole libraries of tracematches are routinely applied to software as it is developed. For example, libraries can be associated with a collection of tracematches specifying constraints on how the library should be used. In these cases we expect that only some tracematches will actually apply to the program under analysis, and the quick check is a sound and simple way to eliminate those that don't apply. We expect this check to become a standard part of the `abc` compiler and it will be enabled by default at the -O level of optimization.

The second stage, flow-insensitive analysis to find consistent shadows, was also effective in some cases, and is also not a very complex analysis once one has a good points-to analysis available. We did find that a context-sensitive points-to analysis was necessary and this turned out to be an ideal use case for demand-driven context-sensitive analysis since we were only interested in the points-to information of variables involved in shadows. Based on our results, we think that this consistent-shadows analysis should be available at a higher-level optimization level (-O3), to be used when run-time overheads are high. In many cases we expect the overheads of a program optimized that way to be lower than those of programs using a hand-coded AspectJ monitor which is not analyzable.

Although we expected that the third stage, the flow-sensitive active-shadows analysis, would give us even more improvement, we found that it did not. To implement and test this analysis we developed a lot of machinery to represent the inter-procedural abstraction of the matching automata, and techniques to soundly approximate even in the presence of threads. To our surprise, the end result is that this extra machinery did not lead to more precise shadow removal. However, this exercise did provide an analysis basis and some new insight into the problem and we think that further refinements to this approach are worth further investigations. We plan to work on this by experimenting with new kinds of must and hybrid points-to abstractions and by improving upon the treatment of multi-threading, perhaps by using the *May Happen in Parallel* (MHP) analysis which is currently being integrated into Soot [19].

Acknowledgements We are very grateful to Manu Sridharan for providing his points-to analysis to us and helping us integrate it into Soot. Steve Blackburn and Robin Garner helped with setting up the DaCapo benchmark suite and interpreting our benchmark results. Stephen Fink helped to clarify our interpretation of his related work. Nomair Naeem contributed with clarifying discussions and by commenting on a draft of this article. We also wish to thank the whole development group of the abc compiler, and in particular the tracematch guys, Pavel Avgustinov and Julian Tibble, for their enduring support and motivation.

References

[1] Chris Allan, Pavel Avgustinov, Aske Simon Christensen, Laurie Hendren, Sascha Kuzins, Ondřej Lhoták, Oege de Moor, Damien Sereni, Ganesh Sittampalam, and Julian Tibble. Adding Trace Matching with Free Variables to AspectJ. In *Object-Oriented Programming, Systems, Languages and Applications*, pages 345–364. ACM Press, 2005.

[2] AspectJ Eclipse Home. The AspectJ home page. http://eclipse.org/aspectj/, 2003.

[3] Pavel Avgustinov, Aske Simon Christensen, Laurie Hendren, Sascha Kuzins, Jennifer Lhoták, Ondřej Lhoták, Oege de Moor, Damien Sereni, Ganesh Sittampalam, and Julian Tibble. *abc*: An extensible AspectJ compiler. In *Aspect-Oriented Software Development (AOSD)*, pages 87–98. ACM Press, 2005.

[4] Pavel Avgustinov, Aske Simon Christensen, Laurie Hendren, Sascha Kuzins, Jennifer Lhoták, Ondřej Lhoták, Oege de Moor, Damien Sereni, Ganesh Sittampalam, and Julian Tibble. Optimising AspectJ. In *Programming Language Design and Implementation (PLDI)*, pages 117–128. ACM Press, 2005.

[5] Pavel Avgustinov, Julian Tibble, Eric Bodden, Ondřej Lhoták, Laurie Hendren, Oege de Moor, Neil Ongkingco, and Ganesh Sittampalam. Efficient trace monitoring. Technical Report abc-2006-1, http://www.aspectbench.org/, 03 2006.

[6] Pavel Avgustinov, Julian Tibble, and Oege de Moor. Making trace monitors feasible. Technical Report abc-2007-1, http://www.aspectbench.org/, 03 2007.

[7] S. M. Blackburn, R. Garner, C. Hoffman, A. M. Khan, K. S. McKinley, R. Bentzur, A. Diwan, D. Feinberg, D. Frampton, S. Z. Guyer, M. Hirzel, A. Hosking, M. Jump, H. Lee, J. E. B. Moss, A. Phansalkar, D. Stefanović, T. VanDrunen, D. von Dincklage, and B. Wiedermann. The DaCapo benchmarks: Java benchmarking development and analysis. In *OOPSLA '06: Proceedings of the 21st annual ACM SIGPLAN conference on Object-Oriented Programing, Systems, Languages, and Applications*, pages 169–190, New York, NY, USA, October 2006. ACM Press.

[8] Feng Chen and Grigore Rosu. Java-MOP: A Monitoring Oriented Programming Environment for Java. In Nicolas Halbwachs and Lenore D. Zuck, editors, *TACAS*, volume 3440 of *Lecture Notes in Computer Science*, pages 546–550. Springer, 2005.

[9] Marcelo d'Amorim and Klaus Havelund. Event-based runtime verification of Java programs. In *WODA '05: Proceedings of the third international workshop on Dynamic analysis*, pages 1–7, New York, NY, USA, 2005. ACM Press.

[10] Maryam Emami, Rakesh Ghiya, and Laurie J. Hendren. Context-sensitive interprocedural points-to analysis in the presence of function pointers. In *PLDI '94: Proceedings of the ACM SIGPLAN 1994 conference on Programming language design and implementation*, pages 242–256, New York, NY, USA, 1994. ACM Press.

[11] Stephen Fink, Eran Yahav, Nurit Dor, G. Ramalingam, and Emmanuel Geay. Effective typestate verification in the presence of aliasing. In *ISSTA'06: Proceedings of the 2006 international symposium on Software testing and analysis*, pages 133–144, New York, NY, USA, 2006. ACM Press.

[12] Allen Goldberg and Klaus Havelund. Automated runtime verification with Eagle. In Ulrich Ultes-Nitsche, Juan Carlos Augusto, and Joseph Barjis, editors, *Workshop on Verification and Validation of Enterprise Information Systems (VVEIS)*. INSTICC Press, 2005.

[13] Erik Hilsdale and Jim Hugunin. Advice weaving in AspectJ. In *AOSD '04: Proceedings of the 3rd international conference on Aspect-oriented software development*, pages 26–35, New York, NY, USA, 2004. ACM Press.

[14] Michael Hind. Pointer analysis: haven't we solved this problem yet? In *PASTE '01: Proceedings of the 2001 ACM SIGPLAN-SIGSOFT workshop on Program analysis for software tools and engineering*, pages 54–61, New York, NY, USA, 2001. ACM Press.

[15] John B. Kam and Jeffrey D. Ullman. Monotone data flow analysis frameworks. *Acta Informatica*, 7:305–317, 1977.

[16] E. Kindler. Safety and liveness properties: A survey. *Bulletin of the European Association for Theoretical Computer Science*, 53:268–272, 1994.

[17] Patrick Lam, Viktor Kuncak, and Martin Rinard. Generalized typestate checking using set interfaces and pluggable analyses. *SIGPLAN Not.*, 39(3):46–55, 2004.

[18] Ondřej Lhoták and Laurie Hendren. Scaling Java points-to analysis using Spark. In G. Hedin, editor, *Compiler Construction, 12th International Conference*, volume 2622 of *LNCS*, pages 153–169, Warsaw, Poland, April 2003. Springer.

[19] Lin Li and Clark Verbrugge. A Practical MHP Information Analysis for Concurrent Java Programs. In Rudolf Eigenmann, Zhiyuan Li, and Samuel P. Midkiff, editors, *LCPC*, volume 3602 of *Lecture Notes in Computer Science*, pages 194–208. Springer, 2004.

[20] Michael Martin, Benjamin Livshits, and Monica S. Lam. Finding application errors using PQL: a program query language. In *Proceedings of the 20th Annual ACM SIGPLAN Conference on Object-Oriented Programming, Systems, Languages and Applications*, pages 365–383, 2005.

[21] Hidehiko Masuhara, Gregor Kiczales, and Christopher Dutchyn. A compilation and optimization model for aspect-oriented programs. In Görel Hedin, editor, *CC*, volume 2622 of *Lecture Notes in Computer Science*, pages 46–60. Springer, 2003.

[22] Klaus Ostermann, Mira Mezini, and Christoph Bockisch. Expressive pointcuts for increased modularity. In Andrew P. Black, editor, *ECOOP*, volume 3586 of *Lecture Notes in Computer Science*, pages 214–240. Springer, 2005.

[23] Manu Sridharan and Rastislav Bodík. Refinement-based context-sensitive points-to analysis for Java. In *PLDI '06: Proceedings of the 2006 ACM SIGPLAN conference on Programming language design and implementation*, pages 387–400, New York, NY, USA, 2006. ACM Press.

[24] Volker Stolz. Temporal assertions with parametrised propositions. In *Seventh Workshop on Runtime Verification, Vancouver, Canada*, March 2007. To appear in *Lecture Notes of Computer Science.*

[25] Volker Stolz and Eric Bodden. Temporal Assertions using AspectJ. *Electronic Notes in Theoretical Computer Science*, 144(4):109–124, 2006.

[26] Robert E. Strom and Shaula Yemini. Typestate: A programming language concept for enhancing software reliability. *IEEE Transactions on Software Engineering*, 12(1):157–171, 1986.

[27] Ken Thompson. Programming techniques: Regular expression search algorithm. *Communications of the ACM*, 11(6):419–422, 1968.

[28] Raja Vallée-Rai, Phong Co, Etienne Gagnon, Laurie Hendren, Patrick Lam, and Vijay Sundaresan. Soot - a java bytecode optimization framework. In *CASCON '99: Proceedings of the 1999 conference of the Centre for Advanced Studies on Collaborative research*, page 13. IBM Press, 1999.

[29] Robert Walker and Kevin Viggers. Implementing protocols via declarative event patterns. In *ACM Sigsoft International Symposium on Foundations of Software Engineering (FSE-12)*, pages 159–169, 2004.