Algorithms for Infeasible Path Calculation

Jan Gustafsson, Andreas Ermedahl, and Björn Lisper
Department of Computer Science and Electronics, Mälardalen University
Box 883, S-721 23 Västerås, Sweden
{jan.gustafsson, andreas.ermedahl, bjorn.lisper}@mdh.se

Abstract
Static Worst-Case Execution Time (WCET) analysis is a technique to derive upper bounds for the execution times of programs. Such bounds are crucial when designing and verifying real-time systems. A key component in static WCET analysis is to derive flow information, such as loop bounds and infeasible paths.

Such flow information can be provided as either as annotations by the user, can be automatically calculated by a flow analysis, or by a combination of both. To make the analysis as simple, automatic and safe as possible, this flow information should be calculated automatically with no or very limited user interaction.

In this paper we present three novel algorithms to calculate infeasible paths. The algorithms are all designed to be simple and efficient, both in terms of generated flow facts and in analysis running time. The algorithms have been implemented and tested for a set of WCET benchmarks programs.

1 Introduction
To give timing guarantees for embedded and real-time systems, a key parameter is the worst-case execution time (WCET) of the software. A static WCET analysis finds an upper bound to the WCET of a program, relying on mathematical models of the software and hardware involved. Given that the models are correct, the analysis will derive a timing estimate that is safe, i.e., greater than or equal to the WCET.

To statically derive a timing bound for a program, information on both the hardware timing characteristics, such as the execution time of individual instructions, as well as the program’s possible execution flows, to bound the number of times the instructions can be executed, needs to be derived. The latter includes information about the maximum number of times loops are iterated, which paths through the program that are feasible, execution frequencies of code parts, etc.

The goal of flow analysis is to calculate such flow information as automatically as possible. Flow analysis research has mostly focused on loop bound analysis, since upper bounds on the number of loop iterations must be known in order to derive WCET estimates [8].

Flow analysis can also identify infeasible paths, i.e., paths which are executable according to the control-flow graph structure, but not feasible when considering the semantics of the program and the possible inputs. Information on infeasible paths is not necessary to find a WCET estimate, but may tighten it.

This article presents ongoing work to automatically calculate infeasible paths. Three new and complementary algorithms are presented. They have been implemented in our prototype WCET analysis tool and tested for a set of WCET benchmarks programs.

The concrete contributions of this article are:
• We present ongoing work to extend our flow analysis method, called abstract execution, to calculate information about infeasible paths.
• We present three algorithms, calculating different types of infeasible path information, allowing us to trade analysis time for flow information precision.
• We show how to make our infeasible path algorithms input data dependent, allowing us to calculate more precise flow information for a program with limitations on its possible input data values.
• We evaluate the effect of our different infeasible path detection algorithms, including the type and amount of flow information generated.

The rest of this paper is organized as follows: In Section 2, we discuss causes of infeasible paths and describe related work. In Section 3, we describe our research prototype, SWEET. Section 4 describes the different algorithms, and Section 5 presents an illustrating example. Section 6 presents analysis results, and in Section 7 we draw some conclusions and discuss future work.
2 Causes of Infeasible Paths and Related Work

There are two different causes to infeasible paths. The first cause is semantic dependencies that always hold, as illustrated by the following code fragment:

\[
\text{if } (x < 0) \text{ A else B; if } (x > 2) \text{ then C else D}
\]

Here, both true-branches for the if statements are always in conflict, and the corresponding path A-C can never be taken.

A second cause to infeasible paths is due to limitations of input data values. Such limitations can be used to further limit the set of feasible paths. For example, if we know that \(x>5\) when the above code is executed, then we can conclude that the paths A-C, A-D, and B-D are all infeasible i.e., we find more infeasible paths with this additional knowledge.

Recent industrial WCET case-studies [7, 9, 17], have shown that it is important to develop good support for both loop bound analysis and infeasible path detection, thereby reducing the need for manual annotations. The case studies also showed that a mode- (giving a WCET estimate under certain system conditions) and input-sensitive WCET analysis often was preferable, in order to obtain better resource utilization and provide a better understanding of the system’s timing characteristics. Thus, it should be important to develop input-sensitive infeasible path analyses.


The proposed infeasible path detection algorithms all use our flow analysis method abstract execution [10, 11], which is briefly described in the next section. This method has some similarities with the one of Lundqvist and Stenström [15], as well as with trace partitioning [3]. However, abstract execution uses a more detailed value domain, and it is based on an abstract interpretation framework.

3 SWEET and Abstract Execution

SWEET (SWEdish Execution time Tool) is a prototype WCET tool developed at Uppsala and Mälardalen University [16]. It consists of three main parts; a flow analysis which detects program flow constraints, a low-level analysis, where timing for program parts are obtained [6], and a final calculation where the longest execution path is extracted given information derived in the two preceding stages [8].

The current flow analysis of SWEET uses a scope-graph [8]. Each scope in the scope-graph is a different execution environment of a program, such as a function or a loop. See Figure 3 for an example. Our current scope-graph representation is context-sensitive, i.e., each call to a function or a loop in a function generates a different scope. Different calls to a function are analysed separately, which may yield higher precision but also a costlier analysis.

Abstract execution is a form of symbolic execution [10, 11], which is based on abstract interpretation. Rather than using traditional fixed-point iteration [5], abstract execution executes the program in the abstract domain, with abstract values for the program variables, and abstract versions of the operators in the language. For instance, the abstract domain can be the domain of intervals: each numeric variable will then hold an interval rather than a number, and each assignment will calculate a new interval from the current intervals held by the variables. As usual in abstract interpretation, the abstract value held by a variable, at some point, represents a set containing the actual concrete values that the variable can hold at that point.

With abstract values, conditionals cannot always be decided, and the abstract execution must then execute both branches. In order to curb the growing number of paths, merging of abstract values for different paths can take place. A merged abstract value then surely contains all the possible concrete values from both paths, and a single-path abstract execution, representing the execution of both paths, can continue from the merging point. Typical merge points are places where different program flows meet, like after if-statements or loops. Merging may yield abstract values that represent the possible set of concrete values in a less precise way: for instance, the merge of the intervals [6..6] and [10..11] is [6..11], which also contains the concrete values 7, 8, 9 not present in the original intervals.

SWEET currently supports abstract execution with intervals. It allows the user to control the placement of merge points, in order to explore different tradeoffs.
between analysis speed and precision. Currently, the
user can specify merge points to be one or more of the
following types: after if-statements, after loop bodies,
after loop exits and after function exits.

Figure 1 gives a simple example of abstract execution
with intervals. The loop in Figure 1(a) is ab-


detected infeasible paths. They are input-sensitive,
and a corresponding “infeasible node flow fact” can be


generated. An example is:

```
scope : <> : #BB82 = 0;
```

specifying that basic block BB82 is not executed in any
iteration of the scope scope.

4.2 Detecting Infeasible Pair of Nodes

The second algorithm finds infeasible pairs of ba-
sic blocks, i.e., blocks which are always excluding each
other during the same iteration of a scope. This gives
additional knowledge as compared to the first analysis,
since there might be nodes which both can be executed
during some iteration of a scope, but which never can
be executed together. The limitation is that infeasible
paths with more than two selections can be missed.

The recorder object for this algorithm is a path (list
of nodes) taken during an iteration. To limit the num-
ber of recorded nodes, only nodes after conditional
branches are recorded. At the entry of a scope or at a
new loop iteration, the path is emptied. Whenever a
conditional branch is taken, we remember the branch

4 Algorithms for Infeasible Paths

We now present our three algorithms for infeasible
path detection. Since they are based on abstract exe-
cution, which is input-sensitive, the analyses are input-
sensitive as well.

All three algorithms have a similar overall structure.
They augment each analysis state with a recorder keep-
ing track of nodes and path(s) taken during a particular
analysis of a scope. Each algorithm resets the recorder
of a scope when starting a new iteration of the scope.
They also associate a collector to each scope, which ac-
cumulates information about nodes and paths during
iterations of the scope. In the end, each collector is
used to generate flow facts for its scope.

4.1 Detecting Infeasible Nodes

The first algorithm finds infeasible nodes, that is:

basic blocks which are never visited in any execution
of a certain scope. Since there is one scope per context,
the resulting flow information becomes context sensi-
tive. An infeasible node is therefore not necessarily
the same as dead code, since the basic block potentially
can be executed in another context.

The recorder object is a bit array with one bit per
node in the scope. These bits are all reset to zero at
each iteration of the scope, and the bit of a node is set
to one at each abstract execution of the node. Thus, a
value of zero, after an iteration, means “definitely not
executed in this iteration” and one means “may have
been executed in this iteration”.

The collector object is a similar bit array. Its bits
are all initialized to zero, and the end of each iteration
the new value of the collector object is set to the bitwise
or of its old value and the current value of the recorder
object. At termination, if the collector holds a zero for
a node, then it is surely never executed in that scope,
and a corresponding “infeasible node flow fact” can be
generated. An example is:

```
scope : <> : #BB82 = 0;
```

specifying that basic block BB82 is not executed in any
iteration of the scope scope.
by appending the corresponding node to end of the path. If both paths are taken, the analysis proceeds in two abstract states, one for each path.

The collector object is a triangular matrix which holds exclusion data. It is of size $N \times N$, where $N$ is the number of possible branch outcomes (basic blocks) for the selections in the scope. The matrix can be triangular since the order of the elements in a pair is irrelevant. All elements in such a matrix are set to $\perp$ in the beginning of the analysis, which means that no information is available to start with. A recorder list $RL$ is added to the collector matrix $M$ when an abstract state has reached the end of a loop body or a function scope. The collector is updated as follows:

\begin{verbatim}
for each node $n_1$ in $RL$ do
  for each subsequent node $n_2$ to $n_1$ in $RL$ do
    $M[n_1,n_2] := 1$
  for each alternative branch node $n_3$ to $n_2$ do
    if $M[n_1,n_3] = \perp$ then
      $M[n_1,n_3] := 0$
    else
      $M[n_1,n_3] := M[n_1,n_3] \lor 0$
\end{verbatim}

For example, if the path $A-C$ was taken in the example in Section 2 we would have updated $M[A][C]$ to 1 and $M[A][D]$ with 0.

When the analysis has finished, the resulting collector matrix is investigated. Matrix positions with $\perp$ mark pairs which have not been touched during the analysis. Some of them can never be executed together anyway due to the structure of the control graph, while the rest really are infeasible pairs. For the first type, generating flow facts will be superfluous. They could be identified using a reachability analysis. However, this is not included in the current implementation, so to avoid a large number of superfluous flow facts, no flow facts are currently generated for $\perp$ positions.

If the matrix positions holds a 0, it marks a node pair that we surely know excludes each other for any iteration of the scope, so for this pair an “excluding pair flow fact” can be generated, like:

\begin{verbatim}
scope : <> : (#BB33 + #BB57) < 2;
\end{verbatim}

specifying that for any iteration of scope the basic blocks BB33 and BB57 are never executed together.

### 4.3 Detecting Infeasible Paths

The third algorithm finds sequences of nodes which are never executed together during the same iteration of a scope. The algorithm makes use of the fact that many infeasible paths can be efficiently represented by allowing them to share a common prefix (sub)path.

The recorder data object is now a tree where each tree node represents a path and has an associated boolean specifying if the corresponding path is feasible or unfeasible. Similar to the recording in the second algorithm we only keep track of nodes taken after branches. However, the tree additionally keeps track of branch outcomes not taken.

Figure 2 gives an illustration of how the recorder tree works. Figure 2(a) gives a CFG with $2^3 = 8$ structurally possible execution paths. Figure 2(b) gives the tree resulting from an execution taking the path $A-B-D-F-G-H-J$ through the CFG. In the recorded tree the paths $A-C$, $A-B-D-E$ and $A-B-D-F-G-I$ have been marked as infeasible. Note that we, for the sake of efficiency, do not record any join nodes in the tree. Similarly, Figure 2(c) gives the tree resulting from an execution path of $A-B-D-E-G-H-J$. Note that the path $A-C$ actually represents $2^2 = 4$ number of paths through the CFG, since $A-C$ is a prefix of all these paths.

For this algorithm the collector is the tree of paths obtained by merging all recorded trees for the scope. The basic idea of the collector is the same as in the first two algorithms, i.e., only keep infeasible path information which are true for all executions of the scope.

Figure 2(d) gives the collector tree resulting from merging the two trees in Figure 2(b) and Figure 2(c). Both trees has $A-C$ as infeasible, and so does the collector tree. Since $A-B-D-E$ is infeasible in Figure 2(b) but not in Figure 2(c), since it is part of the feasible $A-B-D-E-G-H-J$ path, the collector tree cannot keep the path as infeasible. Instead, all paths (both infeasible and feasible) starting with the path $A-B-D-E$ in Figure 2(c) are added to the collector tree. The infeasible path $A-B-D-F$ in Figure 2(c) is extended similarly. The resulting collector tree in Figure 2(d) marks paths $A-C$, $A-B-D-F-G-I$ and $A-B-D-E-G-I$ as infeasible.

A collector tree $CT$ is updated with a recorder tree $RT$ as follows:

\begin{verbatim}
if $RT$ is the first recorded tree reported
  $CT := RT$
else
  for each infeasible path $i$ in $CT$ do
    if $i$ is prefix to a feasible path in $RT$
      mark path $i$ as feasible in $CT$
      add all paths with prefix $i$ in $RT$ to $CT$
\end{verbatim}
After the analysis we create flow facts for the remaining infeasible paths in the collector tree. An example of such a flow fact is:

```
scope : <> : (#BB33 + #BB57 + #BB82) < 3;
```

specifying that basic blocks BB33, BB57, and BB82 are never executed together for each iteration of `scope`.

5 Example

The example code in Figure 3(a) contains infeasible paths of several types (we assume that neither i, x or y are changed in the excluded code). It will be used to illustrate the algorithms we propose in the paper. The program contains five scopes: `main`, `foo1`, and `foo2` (the two calls to `foo`) and the corresponding loop scopes (`foo1_L` and `foo2_L`) in `foo`, as depicted in Figure 3(b). We can identify the following infeasible nodes, pairs and paths:

1. **Infeasible nodes:**
   - E is an infeasible node in `main`, H and I are infeasible nodes in `foo2_L` (limitations in input data).

2. **Infeasible pairs:**
   - B-C, B-E, and D-E are infeasible pairs in `main` (contradicting conditions).
   - A-E and C-E is an infeasible pair in `main` (limitations in input data).
   - G-I and H-J are infeasible pairs in `foo1_L` and `foo2_L` (contradicting conditions).
   - H-I is an infeasible pair in `foo2_L` (limitations in input data).

3. **Infeasible paths:**
   - A-D-E, B-C-E, B-C-F, and B-D-E are infeasible paths in `main` (contradicting conditions).
   - B-D-E is an infeasible path in `main` (limitations in input data).

We note that infeasibility can be expressed in several ways, e.g., the infeasible pair `B-C` and the infeasible paths `B-C-E` and `B-C-F` exclude the same paths.

6 Evaluation

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
<th>#LC</th>
<th>#S</th>
<th>#L</th>
</tr>
</thead>
<tbody>
<tr>
<td>adpcm</td>
<td>Adaptive pulse code modulation algorithm</td>
<td>879</td>
<td>65</td>
<td>27</td>
</tr>
<tr>
<td>bs</td>
<td>Binary search for the array of 15 integer elements</td>
<td>114</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>bsort100</td>
<td>Bubblesort program</td>
<td>128</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>cnt</td>
<td>Counts non-negative numbers in a matrix</td>
<td>267</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>compress</td>
<td>Compression using lzw</td>
<td>508</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>cover</td>
<td>Program for testing many paths</td>
<td>640</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>crc</td>
<td>Cyclic redundancy check computation on 40 bytes of data</td>
<td>128</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>duff</td>
<td>Using &quot;Duff’s device&quot; to copy 43 byte array</td>
<td>56</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>edn</td>
<td>Finite Impulse Response (FIR) filter calculations</td>
<td>285</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>expint</td>
<td>Series expansion for computing an exponential integral function</td>
<td>157</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>ffft</td>
<td>Fast Discrete Cosine Transform</td>
<td>239</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>fopen</td>
<td>1024-point Fast Fourier Transform using the Cooley-Turkey algorithm</td>
<td>219</td>
<td>52</td>
<td>30</td>
</tr>
<tr>
<td>fibcall</td>
<td>Iterative Fibonacci, used to calculate fib(30)</td>
<td>72</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>fir</td>
<td>Finite impulse response filter (signal processing algorithms) over a 700 items long sample</td>
<td>276</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>insertsort</td>
<td>Insertion sort on a reversed array of size 10</td>
<td>92</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>janne_complex</td>
<td>Nested loop program</td>
<td>64</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>jfctint</td>
<td>Discrete-cosine transformation on 8x8 pixel block</td>
<td>376</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>lcdnum</td>
<td>Read ten values, output half to LCD</td>
<td>64</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>lcdcap</td>
<td>LU decomposition algorithm</td>
<td>147</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>matrix</td>
<td>Matrix multiplication of two 20x20 matrices</td>
<td>163</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>ndes</td>
<td>Complex embedded code. A lot of bit manipulation, shifts, array and matrix calculations</td>
<td>231</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>ns</td>
<td>Search in a multi-dimensional array</td>
<td>535</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>nsichnek</td>
<td>Simulate an extended Petri net. Automatically generated code with more than 250 if-statements</td>
<td>4253</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>spft-exam</td>
<td>Linear equations by LU decomposition</td>
<td>121</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>sqrt</td>
<td>Root computation of quadratic equations</td>
<td>166</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>select</td>
<td>A function to select the Nth largest number in a floating point array</td>
<td>114</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>statement</td>
<td>Automatically generated code</td>
<td>1276</td>
<td>9</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. Benchmark programs used

We have used programs from the Mälardalen WCET Benchmark to test our calculations. Table 1 gives some basic data about the programs (LC = lines of code), number of iteration scopes (#S), and number of (context-dependant) loops (#L). Table 2 shows the results of the different analyses. It shows the following information: Analysis time in seconds for abstract execution with loop bound analysis only (LB), number of found flow facts (#FF), and analysis time (Time) for each of the three algorithms (IN = infeasible nodes, EP = exclusive pairs, IP = infeasible paths). All measurements were performed on a 1.25 MHz PowerPC G4 processor, 1 Gb memory running Mac OS 10.4.6.

We see that we, with a small extra cost, can find infeasible nodes and paths for some of the benchmarks.
should be noted that these results are based on single-path analysis, i.e., using a single input that leads to a single execution path. We expect more infeasible nodes and paths to be found when we analyse the programs with inputs that leads to multi-path analyses.

7 Conclusions and Future Work

We do think that our results are promising, but they are still somewhat preliminary: the benchmarks used so far are limited to single-path programs, and we only count the number of generated flow facts for infeasible paths. The next step is to extend the evaluation to a larger set of benchmarks, using multi-path analysis, and to also investigate the effect of the derived infeasible path information on the WCET estimate. In particular, we want to try out the algorithms on industrial real-time codes.

We also want to investigate tradeoffs between analysis time and WCET estimate precision. One possibility would be to generate flow information for individual iterations of a scope. This could give tighter WCET estimates, at the expense of longer analysis times.

Another possibility is to generate non-context-sensitive flow facts, valid for all different call-sites of a particular function or loop. This will, in general, give less precise WCET estimates, but for a lower analysis cost.

Table 2. Analysis results

7 Conclusions and Future Work

We do think that our results are promising, but they are still somewhat preliminary: the benchmarks used so far are limited to single-path programs, and we only count the number of generated flow facts for infeasible paths. The next step is to extend the evaluation to a larger set of benchmarks, using multi-path analysis, and to also investigate the effect of the derived infeasible path information on the WCET estimate. In particular, we want to try out the algorithms on industrial real-time codes.

We also want to investigate tradeoffs between analysis time and WCET estimate precision. One possibility would be to generate flow information for individual iterations of a scope. This could give tighter WCET estimates, at the expense of longer analysis times. Another possibility is to generate non-context-sensitive flow facts, valid for all different call-sites of a particular function or loop. This will, in general, give less precise WCET estimates, but for a lower analysis cost.

References


