ABSTRACT

Embedded systems often have real-time constraints. Traditional timing analysis statically determines the maximum execution time of a task or a program in a real-time system. These systems typically depend on the worst-case execution time of tasks in order to make static scheduling decisions so that tasks can meet their deadlines. Static determination of worst-case execution times imposes numerous restrictions on real-time programs, which include that the maximum number of iterations of each loop must be known statically. These restrictions can significantly limit the class of programs that would be suitable for a real-time embedded system. This paper describes work-in-progress that uses static timing analysis to aid in making dynamic scheduling decisions. For instance, different algorithms with varying levels of accuracy may be selected based on the algorithm’s predicted worst-case execution time and the time allotted for the task. We represent the worst-case execution time of a function or a loop as a formula, where the unknown values affecting the execution time are parameterized. This parametric timing analysis produces formulas that can then be quickly evaluated at run-time so dynamic scheduling decisions can be made with little overhead. Benefits of this work include expanding the class of applications that can be used in a real-time system, improving the accuracy of dynamic scheduling decisions, and more effective utilization of system resources.

1. INTRODUCTION

Software is being embedded as a component of an increasing number of critical systems. Often embedded systems have timing constraints that must be met or the system is not considered functional. Computations that are less precise and on-time are often considered better than more precise computations that are late. This type of system is referred to as real-time. In order to verify that real-time systems will meet their deadlines, designers require that the worst-case execution time (WCET) of each task in a real-time system be known. The process of automatically and statically determining the WCET of a program or task is called timing analysis. Scheduling decisions are based on each task’s WCET and the total time in the schedule. Many requirements are often imposed on real-time programs so that the WCET may be predicted. One of these requirements is that the maximum number of iterations of all loops be known statically. Unfortunately, this requirement significantly limits the class of programs that can be performed in a real-time embedded system.

This paper describes how static timing analysis can be used to aid in making dynamic scheduling decisions. The WCET of a function or a loop is represented as a formula, where the values affecting the execution time are parameterized. Such formulas can then be quickly evaluated at run-time so dynamic scheduling decisions can be made when scheduling a task or choosing algorithms within a task. Benefits of this parametric timing analysis include expanding the class of applications that can be used in a real-time system, improving the accuracy of dynamic scheduling decisions, and more effective utilization of system resources.

2. RELATED WORK

Recently, a number of research groups have addressed various issues in the area of predicting the WCET of real-time programs. Conventional methods for static analysis have been extended from unoptimized programs on simple CISC processors [PuK89, Par93, HW92] to optimized programs on pipelined RISC processors [ZBN93, LBJ94, WH95], from uncached architectures to instruction caches [AMW94, LMW95, HBL95] and data caches [KMH96, LMW96, WMH97], and from assuming that all paths are feasible to automatically detecting constraints on paths [HeW99, EnE00]. All of these methods obtain discrete values to bound the WCET in a non-parametric fashion. Chapman et al. [CBW96] used path expressions to combine a source-oriented parametric approach of WCET analysis with timing annotations, verifying the latter through the former. Bernat et al. [BeB00] also proposed using algebraic expressions to represent the WCET of subprograms, where the algebraic expression is parameterized by some of the subprogram’s parameters. These approaches are closely related to our work in the sense that we also utilize a parametric approach for WCET prediction. However, our approach differs in a number of ways. Instead of a source-level analysis that cannot easily take architectural details of a processor into account, we perform our analysis within an existing framework of tools operating at a low-level intermediate code, which is equivalent to machine code. The other approaches rely on a computational algebra system, like Mathematica or Maple, to evaluate the algebraic expression representing the WCET with the goal of making more accurate predictions. However, relying on a general computational algebra system to calculate the WCET to make dynamic scheduling decisions in real-time would be unrealistic. We instead emphasize the generation of simple formulas where the WCET can be quickly calculated at run-time. Thus, we can retain the parametric model beyond static timing analysis for a multitude of applications ranging from scheduling of tasks to selection of algorithms within a task.
3. NUMERIC TIMING ANALYSIS

Figure 1 depicts the overall organization of the existing numeric timing analysis environment before it was modified to support parametric analysis. We denote this environment as numeric since the timing analyzer must produce a number representing the WCET of a program (i.e., the number of iterations for each loop must be a number that is known to the timing analyzer). An optimizing compiler [BeD88] was modified to produce control flow and branch constraint information as a side effect of the compilation of a file [AMW94, HSR98, HeW99]. A static cache simulator uses the control flow information to construct a control-flow graph of the program that consists of the call graph and the control flow of each function. The program control-flow graph is then analyzed and a caching categorization for each instruction in the program is produced [AMW94]. A separate categorization is given for each loop level in which the instructions and data references are contained. Next, a timing analyzer uses the control flow and constraint information, caching categorizations, and machine dependent information (e.g., pipeline characteristics) as input to make timing predictions [AMW94, HWI95]. Given a program’s control-flow information and instruction caching categorizations along with the processor’s instruction set information, the timing analyzer then derives best-case and worst-case estimates for each path, loop, and function within the program. A timing analysis tree is constructed, where each node of the tree corresponds to a loop or function in the function instance graph. Each node is considered a natural loop, where each function is treated as a loop executing a single iteration. The nodes in the timing analysis tree are processed in a bottom-up manner. In other words, the WCET and BCET for a node are not calculated until the times for all of its immediate child nodes are known. This means that the timing analyzer determines the execution time for programs by first analyzing the innermost loops and functions, and proceeding to higher level loops and functions until it reaches main(). Finally, a graphical user interface is invoked that allows a user to request predictions for portions of the program [KHR96].

Our numeric timing analyzer requires that the number of iterations of each loop be known, that the total number of iterations for a nonrectangular loop nest can be calculated, or that a user provide this information when it is unknown. Figure 2 depicts an abstraction of the loop analysis algorithm that is used. The algorithm repeatedly selects the longest path through the loop until a fixed point is reached (i.e., the caching behavior does not change and the cycles for the worst-case path is guaranteed to remain the same). The maximum number of times that a worst-case path is selected is the minimum of \( m+1 \) or \( n \), where \( m \) is the number of unique paths through the loop and \( n \) is the number of loop iterations. Note that the WCETs for inner loops are predicted before outer loops, an inner loop is treated as being a single node in a path of the next level outer loop, and the control flow is partitioned if the number of paths within a loop exceeds a specified limit [AlY97].

The correctness of the fixed-point algorithm is discussed in more detail elsewhere [AMW94].

```plaintext
cycles = curr_iter = 0.
DO
    curr_iter += 1.
    Find the longest path wcpath.
    cycles += wcpath->cycles.
    WHILE wcpath’s caching behavior can change
        base_cycles = cycles - (wcpath->cycles * curr_iter).
        cycles += wcpath->cycles * (n - curr_iter).
    END WHILE
END DO
```

Figure 2: Numeric Loop Analysis Algorithm

4. PARAMETRIC TIMING ANALYSIS

It would be desirable to support parametric timing predictions when the number of iterations for a loop is unknown until run time. Our goal is to calculate a formula (or closed form) for the WCET for such a loop, where the formula depends on \( n \), the number of iterations for that loop. This formula needs to be relatively inexpensive to calculate since it will be used at run-time to make dynamic decisions concerning the selection of tasks and/or the selection of algorithms within a task. Rather than always passing a numeric value representing the number of cycles of loops and functions up the timing analysis tree, we can instead pass up a symbolic formula when the number of iterations is unknown in a loop representing a node in the subtree.

Figure 3 shows an abstraction of the revised loop analysis algorithm. The algorithm iterates until a fixed point is reached where the caching behavior of the WCET path does not change. The base cycles that are applied before the final worst-case path time is known are saved. Equation 1 shows that the WCET of a loop depends on these base cycles and the WCET path time, which are both constants, and the number of iterations, which is unknown until run time.

```plaintext
WCET_loop_time = WCET_path_time * n + base_cycles (1)
```

If the actual number of iterations exceeds the number of iterations that are required to reach the point where the parametric WCET loop time is calculated, then the parametric result should be comparable to that produced by the numeric timing analyzer when there is a known number of iterations. If the actual number of iterations is less than the number of iterations that are required to reach this point, then there could be an overestimation. For instance, consider the situation as depicted in Table 1. The formula representing the parametric WCET_loop_time would 11\(n+16 \) since the WCET_path_time after the caching behavior does not change is 11 cycles and the base cycles is 16 (20-11 + 18-11). If the loop iterates only once, then the predicted WCET would be 27 cycles, while the actual WCET would be 20 cycles. We could have modified the formula to address dealing with the special cases for having fewer iterations. However, the resulting formula would be more complex and the additional cost of

![Figure 1: Overview of Timing Analysis Process](image)
making these checks at run time would likely outweigh the benefit obtained for a more accurate prediction. Furthermore, if the number of iterations is small, then the loop will likely have little impact on the total WCET. If the number of iterations is large, then there should be very little sacrifice in the accuracy of the predictions.

\[
\text{WCET}_{\text{loop time}} = \max(\text{WCET}_{\text{path A time}}, \text{WCET}_{\text{path B time}}) \times n + \text{base_cycles} \tag{2}
\]

We have developed a preliminary implementation of the parametric timing analyzer. As in the numeric timing analyzer, certain restrictions are imposed. First, recursive programs are not allowed since cycles in the call graph would complicate the generation of unique function instances. Second, indirect calls are not allowed since an explicit call graph must be generated statically. Finally, loops must be structured, which means there must be only a single entry point in the loop. Unlike the numeric timing analyzer, the parametric timing analyzer does not need to know a constant value for the maximum number of iterations of each loop. However, the compiler must be able to generate a symbolic expression that represents the number of loop iterations to be executed at run-time.

Table 2 describes a small set of test programs that were used to contrast the results with numeric and parametric timing analysis. These test programs were all relatively simple since this preliminary implementation does not yet address the case where the worst-case path within a loop cannot be selected due to two or more of the loop paths having a symbolic WCET.

![Figure 4: Example of an Outer Loop with Multiple Paths](image)

Table 1: WCET Path Information for a Hypothetical Loop

<table>
<thead>
<tr>
<th>How the Path Is Evaluated</th>
<th>Path 1</th>
<th>Path 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Execution Including Cache Misses</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>After Caching Behavior Does Not Change</td>
<td>11</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 2: Test Programs

<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matcnt</td>
<td>Counts and sums the nonnegative elements of an integer matrix.</td>
</tr>
<tr>
<td>Matmul</td>
<td>Multiplies two integer matrices and stores the result in a third integer matrix.</td>
</tr>
<tr>
<td>Stats</td>
<td>Calculates the sum, mean, variance, and standard deviation and the linear correlation coefficient between two vectors of numbers.</td>
</tr>
<tr>
<td>Summinmax</td>
<td>Sum the minimum and maximum of the corresponding elements of two integer vectors.</td>
</tr>
<tr>
<td>Sunanegpos</td>
<td>Sums the negative, positive, and all elements of an integer vector.</td>
</tr>
</tbody>
</table>
parametric timing analyzer. The results show that the parametric analysis is less accurate when loops have a single iteration. This is due to the caching behavior of the WCET path not reaching a fixed point, as described in Figure 3. However, the parametric analysis is almost as accurate as the numeric analysis when the number of iterations increases.

5. USING PARAMETRIC WCET PREDICTIONS

One may ask how can a symbolic formula representing a parametric WCET for a loop or function actually be used. A scheduler can dynamically adjust priorities based on a parametric estimate of the WCET of a task. In addition, it would be beneficial to make dynamic decisions within a real-time task based on an estimate of the remaining execution time. Multiple versions of algorithms could be developed that solve the same problem with different degrees of precision and execution time. Parametric WCET prediction can potentially allow the selection of an algorithm with higher precision due to tighter WCET predictions. Likewise, parametric WCET analysis may be used to determine an appropriate number of iterations for algorithms that progressively improve the precision of calculations (i.e., imprecise computations). The functions providing the parametric WCET can be provided as call-backs to the operating system. Once the parameters, such as the number of loop iterations, have been supplied by the application, the scheduler may inspect the time budget to determine how to schedule the remainder of a task’s execution.

If code within a task is generated to represent the symbolic WCET formula during the parametric timing analysis, then it would seem that this code would affect the previously determined timing predictions due to changing the addresses of the instructions and potentially the caching behavior. Rather than inserting the code for the symbolic WCET formula directly into the source code at the point it is used, we decided to invoke a function that evaluates the symbolic WCET formula and appends these functions to the end of the program. We illustrate the approach with an example given in Figure 5, where the timing analysis is accomplished in stages as the parametric formulas are generated and later evaluated. In this example function #4 is generated by the timing analyzer to calculate the WCET of loop 3, whose number of iterations is unknown until run time.

1. A call to a function is inserted that returns the WCET for a specified loop or a function based on a parameter indicating the number of loop iterations that becomes known at run time. The instructions that are associated with the call and that use the return value after the call are generated during the initial compilation. For instance, in Figure 5(a) function #1 calls function #4 to obtain the WCET of loop #3, which contains a symbolic number of iterations.

2. The timing analyzer generates the source code for the called function into a separate file when processing the specified loop or function whose time needs to be calculated at run time. For instance, Figure 5(b) shows that after loop #3 has been parametrically analyzed, the code for function #4 has been generated. Note that the timing analysis tree representing the loops and functions in the program is processed in a bottom up fashion. The code in function 1 invoking the generated function is not evaluated until after the generated function is produced. The static cache simulator can initially assume that a call to an unknown function invalidates the entire cache. Figure 6 shows an example of the source code of such a generated function. This function returns the number of cycles associated with the formula given in Table 3 for the Matcnt program.

3. The generated function is compiled and is placed at the end of the executable. The formula representing the symbolic WCET need not be simplified by the timing analyzer. Most optimizing compilers perform constant folding, strength reduction, and other optimizations that will automatically simplify the symbolic WCET produced by the timing analyzer. By placing the generated function after the rest of the program, the other instruction addresses in the program are unaffected. Likewise, the predicted caching behavior of the previously analyzed functions and loops are also unaffected since we are processing the tree in a bottom-up order.

4. The timing analyzer is reinvoked to resume the analysis of the program, which includes calculating the WCET of the generated function and the code invoking such a function. For instance, Figure 5(c) shows that the generated function #4 has been numerically analyzed and Figure 5(d) shows that function #1 has been parametrically analyzed, which includes the numeric WCET required for executing function #4.

<table>
<thead>
<tr>
<th>Program</th>
<th>Formula</th>
<th>n</th>
<th>Observed Cycles</th>
<th>Numeric Analysis</th>
<th>Parametric Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Estimated Cycles</td>
<td>Ratio</td>
</tr>
<tr>
<td>Matcnt</td>
<td>$160n^2 + 267n + 857$</td>
<td>1</td>
<td>1,206</td>
<td>1,207</td>
<td>1.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>18,954</td>
<td>19,523</td>
<td>1.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1,622,034</td>
<td>1,627,553</td>
<td>1.003</td>
</tr>
<tr>
<td>Matmul</td>
<td>$33n^3 + 310n^2 + 530n + 851$</td>
<td>1</td>
<td>1,432</td>
<td>1,460</td>
<td>1.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>62,182</td>
<td>70,137</td>
<td>1.126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>33,725,782</td>
<td>36,153,837</td>
<td>1.072</td>
</tr>
<tr>
<td>Stats</td>
<td>$1049n + 1959$</td>
<td>1</td>
<td>2,885</td>
<td>2,912</td>
<td>1.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>12,290</td>
<td>12,449</td>
<td>1.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>106,340</td>
<td>106,859</td>
<td>1.005</td>
</tr>
<tr>
<td>Summinmax</td>
<td>$17n + 123$</td>
<td>1</td>
<td>139</td>
<td>140</td>
<td>1.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>283</td>
<td>293</td>
<td>1.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1,723</td>
<td>1,823</td>
<td>1.058</td>
</tr>
<tr>
<td>Snnegpos</td>
<td>$13n + 94$</td>
<td>1</td>
<td>104</td>
<td>107</td>
<td>1.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
<td>203</td>
<td>224</td>
<td>1.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>1,193</td>
<td>1,394</td>
<td>1.168</td>
</tr>
</tbody>
</table>

Table 3: Results
(a) Loop #3 Contains a Symbolic Number of Iterations

(b) Loop #3 Is Analyzed and a WCET Function Is Generated

(c) Generated Function Is Analyzed

(d) Function Containing Code Calling Generated Function Is Analyzed

Figure 5: Example of Using Parametric Timing Predictions

```
int WCET_loop(int n)
{
    return (160*n + 267)*n+857;
}
```

Figure 6: Example of a Generated Function

Calculating the WCET of a Loop

In summary, this approach allows for the timing analysis to be accomplished in stages. Parametric formulas are produced when needed and source code functions representing these formulas are generated. These generated functions are then compiled, analyzed, and revised timing predictions are made. This process continues until a formula can be produced for the entire program or task.

6. FUTURE WORK

There are several types of problems that we have yet to address in our parametric timing analysis. Sometimes the number of iterations executed by a loop can vary since it may depend on the value of an outer loop counter variable. Our approach to resolve this in the numeric timing analysis was to formulate a non-rectangular loop nest in terms of summations and solve the resulting equation [HSR00]. If the range of all of the loop variables in the nonrectangular loop nest cannot be bounded, then the summations can still be formulated and an equation in a closed form can still often be generated when performing parametric timing analysis.

Likewise constraints on branches can limit the paths that can be taken in a program. Automatically detecting these constraints can result in tighter WCET predictions [HeW99]. We need to formulate an equation when such branch constraints are present and the number of loop iterations is only known at run time.

Finally, we need to implement the approach described in Section 5. This will involve modifying the timing analyzer to recognize the names of special function calls indicating that they should return the WCET of functions or loops. We will also have to change the environment shown in Figure 1 to repeatedly invoke the compiler, static simulator, and timing analyzer as new functions are generated and then subsequently analyzed.

7. CONCLUSIONS

We have described how static timing analysis can be used to aid in making dynamic scheduling decisions. We have shown how a symbolic formula can be produced to represent the WCET for a loop when the number of loop iterations cannot be determined until run-time. While more complex cases should be addressed, preliminary results for some simple programs were given indicating that the approach for parametric timing analysis is almost as accurate as a completely numeric approach when the number of iterations in a loop exceeds the number of iterations required to reach a point where the WCET caching behavior is in a steady state. We also outlined how these parametric WCETs could be used without affecting the timing predictions of the loops and functions that were previously analyzed. The benefits of parametric timing analysis include expanding the class of applications that can be used in a real-time system, improving the accuracy of dynamic scheduling decisions, and more effective utilization of system resources.

ACKNOWLEDGEMENTS

This research was supported in part by the Spanish CICYT grant TAP98-0333-C03-01 and the National Science Foundation grants EIA-9806525, CCR-9904943, and EIA-0072043. The work performed by Frank Mueller was under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

8. REFERENCES


1988).


