LLVMTA: An LLVM-Based WCET Analysis Tool

Sebastian Hahn
Saarland University, Saarland Informatics Campus, Saarbrücken, Germany

Michael Jacobs
Saarland University, Saarland Informatics Campus, Saarbrücken, Germany

Nils Hölscher
TU Dortmund University, Germany

Kuan-Hsun Chen
University of Twente, The Netherlands

Jian-Jia Chen
TU Dortmund University, Germany

Jan Reineke
Saarland University, Saarland Informatics Campus, Saarbrücken, Germany

Abstract
We present llvmta, an academic WCET analysis tool based on the LLVM compiler infrastructure. It aims to enable the evaluation of novel WCET analysis approaches in a state-of-the-art analysis framework without dealing with the complexity of modeling real-world hardware architectures. We discuss the main design decisions and interfaces that allow to implement new analysis approaches. Finally, we highlight various existing research projects whose evaluation has been enabled by llvmta.

2012 ACM Subject Classification Software and its engineering → Software performance; Software and its engineering → Automated static analysis; Software and its engineering → Compilers; Computer systems organization → Real-time system architecture

Keywords and phrases WCET analysis, low-level analysis, LLVM

Digital Object Identifier 10.4230/OASIcs.WCET.2022.2

Supplementary Material llvmta and a patched version of LLVM are available as open source for academic research purposes:
Software (Source Code): https://gitlab.cs.uni-saarland.de/reineke/llvmta
Software (Source Code): https://gitlab.cs.uni-saarland.de/reineke/llvm

Funding This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the Transregional Collaborative Research Centre SFB/TR 14 (AVACS); by the Saarbrücken Graduate School of Computer Science, which received funding from the DFG as part of the Excellence Initiative of the German Federal and State Governments; and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 101020415).

Acknowledgements We would like to thank all the students at Saarland University who contributed during the development of llvmta: in particular Claus Faymonville for his extensive contributions in a very early development stage, but also Tina Jung, Darshit Shah, and Tobias Stark.

1 Work underlying this paper was performed between 2014 and 2019 while the author was still working at Saarland University. Since 2019, the author is affiliated with AbsInt Angewandte Informatik GmbH which is not related to work described in this paper.

2 Work underlying this paper was performed between 2014 and 2017 while the author was still working at Saarland University. Since 2018, the author is affiliated with Artengis GmbH which is not related to work described in this paper.
1 Introduction

In this paper, we introduce LLVMTA, an open-source worst-case execution time (WCET) analysis tool developed at Saarland University from 2014 onwards. Our aims with this paper are twofold: First, we want to convey the main design goals of LLVMTA and how these design goals manifest in its current design. Second, to facilitate future work on LLVMTA, we describe the most important interfaces, which need to be implemented to adapt and extend LLVMTA.

At the onset of the work on LLVMTA, four WCET analysis tools were at the disposal of the authors: the commercial WCET analyzer aiT by AbsInt GmbH, and the three academic WCET analysis tools otawa [7], developed in Toulouse, Chronos [49] developed in Singapore, and Heptane [38], developed in Rennes. Our goals at the time were to study WCET analysis for microarchitectures that feature timing anomalies [53, 64] and to explore the potential of compositional timing analysis [37]. In order to evaluate the full potential of the planned approaches, we decided to implement them in a state-of-the-art analysis framework. To the best of our knowledge, the required state-of-the-art analysis features (in particular abstract execution graphs, as discussed in Section 2) were only available in the commercial aiT tool created by Absint. Commercial analysis tools as aiT, however, exhibit a high degree of complexity as they have to support a wide range of real-world hardware platforms and need to scale to large real-world applications under analysis. Thus, we decided to create our own analysis framework from scratch, to enable the rapid prototyping of novel analysis approaches. Due to the use of state-of-the-art analysis techniques, it is still reasonable to judge whether the observed gain in precision and/or efficiency of novel analysis approaches would translate to commercial tools.

Also worth mentioning is the WCET compiler WCC [20] developed in Dortmund. WCC is a compiler focusing on WCET optimizations, implementing its own high- and machine-level intermediate representations, namely ICD-C and ICD-LLIR. WCC uses aiT for its timing analysis and does not provide timing analysis on its own. While LLVMTA is using the LLVM compiler infrastructure it does not perform code transformations on its own.

As part of T-CREST [60], a tool called Platin has been developed in Vienna. Platin also uses aiT to provide low-level timing analysis or, alternatively, an internal analyzer, which, however, does not feature a detailed microarchitectural analysis. Platin can compute loop bounds by both static analysis, using LLVM infrastructure, and by simulating short traces. These loop bounds can then be fed as flow facts to aiT.

We set out to create a new academic WCET analysis tool with these minimal requirements:

- Support of precise and accurate analysis of microarchitectures with timing anomalies using state-of-the-art techniques, in particular abstract execution graphs [72].
- Support of compositional analysis [37] approaches in which the analysis of different timing contributors is performed separately.
- Flexibility to easily replace pipeline or cache models or to add new path constraints.

WCET analysis is challenging for several reasons, including some that we were not particularly interested in, which entails some explicit “non-goals”:

- Our goals has been to study fundamental challenges in WCET analysis, not to support particular (commercial) microarchitectures. We note, however, that given sufficient knowledge of the underlying microarchitectures, it would be possible to support particular microarchitectures within LLVMTA.
- WCET analysis typically applies to binary executables. This entails the challenge of reconstructing the control-flow graph of the code, which is not explicit in the binary. LLVMTA bypasses this challenge by integrating into the LLVM compiler.
The remainder of this paper is structured as follows: We begin by giving a brief overview of the architecture of static WCET analyzers. Next, we present an overview of the architecture of llvmta. Subsequently, we present the usage of the command-line tool llvmta for a small example. Finally, we conclude the paper with a brief discussion of existing applications of llvmta by pointing out future steps in the development of the tool.

2 Standard Architecture of Static WCET Analysis Tools

In this section, we describe the de facto standard architecture underlying static WCET analysis tools today, which in particular underlies llvmta. There are fundamentally different WCET analysis approaches, such as measurement and hybrid WCET analysis, which are out of scope in this discussion.

llvmta and other WCET analyzers operate on a control-flow graph (CFG) representation of the program under analysis. A CFG is a directed graph whose nodes correspond to basic blocks, i.e. straight-line code sequences, and whose edges correspond to possible control flow between these nodes. The CFG of a program is not explicit in the machine code executed on the hardware. Thus the first step of most WCET analysis tools is to reconstruct a CFG from the program binary [75, 76, 45, 23, 67, 8]. In llvmta, the CFG is directly obtained from the compiler generating the binary rather than by reconstructing it.

Given the program’s CFG, the WCET analysis problem can then be decomposed into three subproblems:
1. Deriving constraints that approximate the subset of paths through the control-flow graph that are semantically feasible.
2. Determining the possible execution times of program parts, such as basic blocks, accounting for the timing effects of microarchitectural features such as pipelining and caching.
3. Combining the information from 1. and 2. to derive a bound on the program’s WCET.

The first subproblem depends only on the program’s semantics and is thus independent of the underlying microarchitecture. If the program under analysis contains loops, it is necessary to bound each loop’s maximum number of iterations; otherwise, no WCET bound can possibly be derived. Different loop bound analyses have been described in the literature [31, 52, 30, 74, 15, 19, 58, 6, 10]. Generalizing loop bound analysis, control-flow analysis derives constraints on the possible execution paths through the CFG [46, 16, 5, 42, 68, 59, 61] including loop bounds.

By definition, the second subproblem critically depends on the underlying microarchitecture. Thus the underlying analysis is often called microarchitectural analysis. Traditionally, the output of microarchitectural analysis have been bounds on the timing contributions of the basic blocks [7, 38, 49] of the program. As modern processors employ pipelining, the execution of successive basic blocks may overlap substantially. Thus, it is important not to “pay” for this overlap multiple times, and to analyze basic blocks within the context of their surrounding basic blocks. Different approaches to this end have been proposed [51, 77, 39, 53, 12, 22, 17, 78, 48, 65, 82]. We are unable to give a complete account of these approaches here due to space limitations; instead we focus on a brief description of the approach taken in llvmta.

Microarchitectural analysis in llvmta can be seen as a static cycle-by-cycle simulation of the execution of the program on abstract microarchitectural states, accounting for any microarchitectural component that influences the execution’s timing, such as pipelining (including e.g. forwarding effects), branch prediction, load and store buffers, and caches.
Due to abstraction this static simulation may lack information, e.g. whether an access results in a cache hit or a cache miss, or whether two memory accesses alias, the simulation may have to “split” following multiple successor states, introducing nondeterminism. The output of microarchitectural analysis is an abstract execution graph (AEG) [72] whose nodes correspond to abstract microarchitectural states and whose edges correspond to the passage of processor cycles. An important distinguishing feature of this approach is that the AEG may capture correlations between the timing contributions of different basic blocks, rather than computing a single bound for each basic block. The key to make this approach successful in practice is to find abstractions that strike a good balance between analysis complexity and precision. For caches various compact abstractions have been developed [1, 25, 26, 27, 71, 33, 11, 14, 28, 24, 54, 79, 9, 80] that offer varying degrees of precision depending on the underlying replacement policy [40, 63].

The third and final step of WCET analysis is to combine the information gathered in the first two steps to compute a bound on the program’s WCET. The most popular approach to this path analysis problem is the implicit path enumeration technique (IPET) [50]. The basic idea behind IPET is to solve the path analysis problem via an integer linear program (ILP). Integer variables are introduced for each edge in the AEG, encoding the frequency of taking those edges during an execution. The structure of the AEG imposes linear constraints relating these frequencies, implicitly encoding all possible paths through the AEG. Additional constraints are obtained from control-flow analysis; otherwise unbounded solutions would be possible in the presence of loops. Finally the objective function captures the cost of a given path through the AEG. Maximizing the objective function yields a safe WCET bound provided that the previous analyses capture all possible executions of the program on the microarchitecture. The overall WCET analysis flow is depicted in Figure 1.

For a more detailed discussion of static WCET analysis and related techniques we refer to the survey paper by Wilhelm et al. [81]. The same techniques can be used to safely approximate the number of occurrences of other microarchitectural events [43]. E.g. one can similarly determine a bound on the number of cache misses in any possible execution of the program.
3 LLVMTA Tool Architecture

3.1 High-level Structure

We implemented a low-level analysis tool called LLVMTA, following the scheme sketched in Figure 1. In this section, we provide details on this tool. LLVMTA is based on the LLVM compiler infrastructure [47], and it is hooked into the common LLVM compilation flow as depicted in Figure 2.

Overall Tool Architecture. Given a C program, the compiler frontend clang (https://clang.llvm.org) translates the program into the LLVM intermediate representation. After an optional optimization phase (OPT), the program is further translated to the assembler code (LLC) which results in the final binary after the linking step. Our analyses are implemented on the final assembler representation in the LLVM backend which is the representation closest to the machine level. The timing bound determined by LLVMTA is valid for the resulting binary, i.e., it will change accordingly if the binary changes, e.g. due to different compiler optimizations.

The integration of low-level timing analysis and compilation offers several advantages. First, no control-flow reconstruction of the binary is required because control-flow elements such as functions, basic blocks, and loops are provided by the prior compilation step. Second, the low-level analysis in the backend can make use of (high-level) information obtained at earlier stages and maintained during compilation. On the downside, the analysis requires as input the program to be analyzed in LLVM intermediate representation, and provides timing estimates only for the binary produced by the specific compiler. Commonly, this representation can be obtained using the compiler frontend from the high-level source program, for example given in C. It is conceivable, but it has not been experimentally validated, to apply binary lifters [4, 2, 18] to obtain an intermediate representation directly from binaries. The analysis results would then be valid for binaries obtained by recompiling the intermediate representation. Furthermore, the addresses of the instructions and the static data are only known after the linking step and would have to be fed back to the low-level analysis for sound analysis results. This is currently not implemented, but there are no major technical obstacles to doing so.

LLVMTA low-level analysis. To obtain precise results, we have implemented context-sensitive analysis [70], i.e. the analysis distinguishes different contexts that influence the execution behaviour. As an example, the execution behaviour of the first iteration of a loop usually differs from the behaviour of later iterations because the caches are being filled during the first iteration [55]. To establish a context-sensitive analysis framework, we implemented trace partitioning [56] on the final assembler representation in the LLVM backend. Context sensitivity is achieved by partitioning the set of execution traces according to some predicate on traces. We implemented predicates to discriminate different iterations of a loop, as well as different call sites of a function. The degree of context sensitivity, i.e. the number and size of these predicates, is an analysis parameter.

Based on our context-sensitive analysis framework, we have implemented a value analysis that tracks constant values of registers and memory cells. This value information is used to derive address information for data accesses. Despite the simplicity of the analysis domain, it is sufficient to precisely analyse stack-relative accesses. For accesses to globally defined objects such as global arrays, our tool uses information provided by the compiler to determine the range of possible addresses.
In order to derive loop bounds, we use the LLVM-internal scalar evolution analysis that provides an upper bound on the iteration count of loops in their intermediate representation. Our tool matches loops in the assembler representation to loops in intermediate representation in order to automatically obtain upper loop bounds on the assembler level. Manual loop annotations can be provided by the user for loops with complex iteration patterns. The scalar evolution analysis, originally based on [6] and extended in [10], computes a closed-form expression to describe how the values of variables evolve within a single loop iteration. These expressions are used to derive upper loop bounds, either in the form of numeric values or symbolic expressions w.r.t. the function parameters.

Our tool supports the analysis of different generic hardware platforms rather than proprietary industrial platforms for the ARM and the RISC-V instruction sets. This is sufficient to evaluate the general concepts used in timing analysis and takes significantly less effort to implement. We model textbook pipelines (see [41]) with in-order, strictly in-order [35, 32, 36], and out-of-order execution. The microarchitectural analysis supports scratchpad memories, as well as caches with least-recently-used replacement policy and both write-through and write-back policy. We have implemented must, may, and persistence cache analysis [1, 54, 62]. As background memory, the tool supports fixed-latency memory as well as dynamic random-access memory with a closed-page controller and distributed refreshes.

LLVMTA implements the fast-forwarding technique presented in [44] to increase the performance of the microarchitectural analysis. This optimization exploits the fact that pipelines tend to converge while waiting for memory, i.e. the pipeline cannot advance further until the current memory request is finished. Once converged, the (abstract) state of the pipeline stays the same as long as the memory is busy.

The abstract execution graph produced by the microarchitectural analysis is compressed afterwards. LLVMTA supports two different levels of compression. Either all start and end nodes within a basic block are kept separate to allow for a precise path analysis [72], or the graph is compressed into a single edge per basic block to allow for an efficient path analysis. Our tool supports multiple solvers to solve the ILP formulation resulting from the path analysis, including the commercial tools IBM ILOG CPLEX Optimization Studio (https://www.ibm.com/us-en/marketplace/ibm-ilog-cplex) and Gurobi Optimizer (https://www.gurobi.com) that exhibit the best performance [57].

3.2 Limitations

While offering many advantages such as code reuse and flexibility, the nature of an academic prototype and the tight coupling with a compiler infrastructure also comes with limitations.

LLVMTA operates on the machine-level IR rather than on the binary, which may yield results that are not entirely faithful to the generated machine code for the following reasons. The assembler may break down pseudo-assembly instructions used in the machine-level IR – the level we perform the analysis on – into several machine instructions in the actual binary (especially on RISC-V). The address mapping is only determined after linking, and LLVMTA currently operates on a made-up address mapping. We note that it would be possible to obtain a faithful address mapping from the linker. For the ARM instruction set, predication is supported for branch instructions, but not for arbitrary machine instructions as specified in the instruction set architecture. As mentioned earlier, there are currently no microarchitectural models that correspond to existing commercial hardware designs. Finally, the tool is reasonable fast on the standard WCET benchmarks, but will likely not scale to real-world applications.
3.3 Main Design Interfaces

To reach our goal of flexibility, we use shared interfaces. New low-level analyses can be obtained by implementing these interfaces with new classes (possibly inheriting existing ones) and directing the analysis framework to employ these implementations. The most important interfaces of LLVM allow for:

- static program analysis on machine-level LLVM intermediate representation,
- microarchitectural analysis, in particular pipeline modeling,
- cache analysis, and
- additional path analysis constraints.

3.3.1 Program Analysis at Machine-level Intermediate Representation

The interface class `ContextAwareAnalysisDomain` enables context-sensitive analysis on a control-flow graph with machine-level instructions.

Part of the interface specifies the basic operation on abstract domain values in the spirit of abstract interpretation [13]:

- `isBottom`: does the current abstract value represent the bottom element of the analysis lattice?
- `lessequal` compares the current abstract value with another given one w.r.t. the partial order $\preceq$ of the analysis domain,
- `join` joins a given abstract value into the current abstract value w.r.t. the partial order of the analysis domain. The behaviour should be consistent with `lessequal`.

The second part of interface specifies the transfer behaviour of abstract values while abstractly interpreting the control-flow graph of the program under analysis.

- `transfer` takes the next instruction to analyze, the current analysis context, and, optionally, analysis information of preceding static analyses at this program point. It modified the current abstract value by the effect of the instruction in the specified context.
- `guard` can be used to sharpen the current abstract value by the knowledge of the outcome of a branch instruction (either taken or not taken).
- `enterBasicBlock` models the effect of entering a basic block on the analysis information.

3.3.2 Microarchitectural Analysis

The interface class `MicroArchitecturalState` models the abstract state of the microarchitecture under analysis. Microarchitectural analysis, unlike most program analysis techniques operates at the granularity of processor cycles rather than program instructions. `MicroArchitecturalState` has the following interface:

- The constructor of the class creates the initial microarchitectural state from which the state space exploration starts. This usually represents a state with an empty pipeline and unknown cache contents.
- `cycle` models the behavior of executing the machine for a single cycle. The abstract microarchitectural state is modified in-place. It takes a configurable set of precomputed analysis information, e.g. address information for memory-accessing instructions.
- `isFinal` specifies whether a given instruction has just finished execution in the current microarchitectural state. An instruction is hereby identified by its instruction address and a context. This predicate allows use to map microarchitectural states to instructions in the control-flow graph of the program under analysis. This is mostly a technicality, but influences e.g. at which points during the analysis microarchitectural states can be `joined`.
- `isJoinable` and `join` are used to test whether microarchitectural states can be joined - and do so if it is possible.
The interface class also features helper functions to model common functionality found in any microarchitecture such as the program counter and its evolution during program execution. The cycle behaviour implicitly induces an microarchitectural state graph from the initial state up to states in which the last instruction of the program under analysis finished execution. At cycle granularity, i.e. having one edge per single execution cycle, such graphs are too large to be used in path analysis. In Section 3.3.4 below, we will describe how to obtain a more compact graph where edges are collapsed to describe multiple execution cycles at once.

### 3.3.3 Cache Analysis

An important part of the microarchitecture that needs attention during timing analysis are caches. The interface class `AbstractCache` models the abstract cache state of the cache under analysis and specifies the behaviour of the cache replacement policy.

- `update` models the effect of loading from or storing to a given abstract address, usually an address interval.
- `lessequal` and `join` specify the abstract analysis domain by providing basic lattice operations $\sqsubseteq$ and $\sqcup$.

For cache analysis with local classifications [54], the following functions are also relevant:

- `classify` tells for a given abstract address whether an access is guaranteed to hit or miss the cache.

For persistence cache analysis [62], the following functions are also relevant:

- `enterScope` and `leaveScope` model the behaviour on entering and leaving a persistence scope.
- `getPersistentScopes` determines for a given abstract address the scopes in which the address is persistent.

### 3.3.4 Path Analysis

To perform path analysis on the results of microarchitectural analysis, there are mostly two tasks to perform. The first is to determine a compact microarchitectural state graph with what we call `weights` for each edge – collapsing multiple execution cycles at once. Weights can be as simple as numeric values such as the number of cycles, the number of cache misses, etc., but also more complex things such as the set of persistent cache misses. The second task is, to use these weights to build up the constraints of an integer linear program that encodes the worst-case path through the microarchitectural state graph. While solutions to the second task are rather specific to the constraints to be generated, solutions to the first task share sufficient commonalities to be captured by an interface class.

During construction of the microarchitectural state graph, we traverse the implicit graph at cycle granularity as described in Section 3.3.2. To keep the graph compact, LLVMTA joins nodes, i.e. microarchitectural states, and edges where possible (see `isJoinable`) without losing too much precision. The general interface class to construct the weights on these collapsed edges is `StateGraphEdgeWeightProvider`. There is a simpler but more limited interface class for numeric weights only, called `StateGraphNumericEdgeWeightProvider`.

- `extractWeight` takes a part of a microarchitectural state called `LocalMetrics` and extracts a weight (of any type) from it. The LocalMetrics can accumulate any weight within a state during microarchitectural analysis, e.g. the number of cycles executed or the number of cache misses since entering a basic block.
- `joinWeight` combines weights of to-be-collapsed edges with same source and same target node, e.g. by taking the maximum.
Figure 3 Simple AEG visualized with yComp.

- `concatWeight` concatenates two weights of two consecutive to-be-collapsed edges, e.g. by adding two numeric weights.
- `getNeutralWeight` returns the neutral element w.r.t. operation `concatWeight`.

4 Using LLVM-Compilation

In this section, we shortly discuss the usage and the expected output of LLVM-Compilation.

As input, a program written in C is provided to LLVM-Compilation. CLANG translates the C program to its machine intermediate representation that is specific to the selected target architecture (ARM or RISC-V). The `runtests` script takes care of all necessary compilation steps and is the main script to perform timing analysis using LLVM-Compilation. LLVM-Compilation can automatically derive loop bounds using LLVM’s scalar evolution analysis in many cases. If LLVM-Compilation fails to obtain a loop bound, the user has to provide loop bounds manually in a CSV format `LoopAnnotations.csv`. The file can be auto-generated with placeholders to fill the bounds using the `-ta-output-unknown-loops` option.
Table 1: Output files generated by LLVMTA.

<table>
<thead>
<tr>
<th>Compiler Frontend</th>
<th>Assembler (.txt)</th>
<th>Program in LLVM’s machine intermediate representation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assembler (.S)</td>
<td>Unlinked program assembly</td>
</tr>
<tr>
<td>Preanalysis</td>
<td>AnnotatedHeuristics</td>
<td>Contains inserted partitioning directives guiding context-sensitive analysis</td>
</tr>
<tr>
<td></td>
<td>PersistenceScopes</td>
<td>Start and end of all persistence scopes</td>
</tr>
<tr>
<td></td>
<td>CallGraph</td>
<td>List of all statically known callers and callees</td>
</tr>
<tr>
<td></td>
<td>ConstantValueAnalysis</td>
<td>Constant values determined for machine registers and memory cells for each instruction</td>
</tr>
<tr>
<td></td>
<td>LoopBounds</td>
<td>Contains loop bounds determined</td>
</tr>
<tr>
<td></td>
<td>AddressInformation</td>
<td>Addresses of instructions and addresses of data accessed by memory operations</td>
</tr>
<tr>
<td>Microarch. Analysis</td>
<td>MicroArchAnalysis</td>
<td>Invariant set of microarchitectural state at the beginning and end of each basic block</td>
</tr>
<tr>
<td></td>
<td>StateGraph_Time (.vcg)</td>
<td>Microarchitectural state graph</td>
</tr>
<tr>
<td>Path Analysis</td>
<td>LongestPath</td>
<td>ILP path analysis formulation</td>
</tr>
<tr>
<td></td>
<td>PathAnalysis_&lt;Weight&gt;_….&lt;Max</td>
<td>Min&gt;</td>
</tr>
<tr>
<td>Results</td>
<td>TotalBound (.xml)</td>
<td>Machine readable output of calculated bounds</td>
</tr>
<tr>
<td></td>
<td>Statistics</td>
<td>Resource consumption of analysis</td>
</tr>
</tbody>
</table>

During the analysis execution, intermediate results are printed as the different analysis stages are performed. The output files are listed in Table 1. These files help the user understand what happens during the analysis and to inspect intermediate and final results. One important output is the file StateGraph_Time, which contains the abstract execution graph with microarchitectural states as nodes and edges with weights such as timing or number of cache misses. The graph can be viewed with the yComp graph viewer [66], as shown in Figure 3 for the “simplewhile” example, which is provided along with the test suite of LLVMTA. The example consists of a single while loop, incrementing a single variable 50 times before terminating. The WCEP is highlighted by the red edges. It is worth noting that multiple abstract microarchitectural states are created for the while loop in the main function allowing for higher analysis precision than a simple “single execution time per basic block” path analysis scheme.

5 Existing Research Applications of LLVMTA

In this section, we briefly summarize research carried out with the help of LLVMTA.

**Cache Analysis.** Classically, write-back caches have not been used in hard real-time systems as it was not known how to model them in a sufficiently precise way during WCET analysis. This gap has been closed by combining two perspectives: a store-focused one, which answers whether a store may dirtify a clean cache line, and an eviction-focused one, which answers whether a cache miss may evict a dirty cache line and thus cause a write back [9].

We also employed LLVMTA in the development and evaluation of exact cache analysis, in particular of exact cache persistence analyses [73].
Compositional Analysis. LLVM supports compositional analysis approaches, i.e. several weights can be chosen for maximization such as useful cache blocks or accesses to the shared bus. The impact of such independent maximizations of different metrics on the efficiency and precision of WCET analysis has been explored in a master’s thesis [21]. In addition, LLVM can sample interference response curves and calculate compositional base bounds based on the results of a microarchitectural analysis that safely covers all possible cases of temporal interference [34]. In this way, it bridges the gap between schedulability analyses, which typically rely on timing compositionality, and modern microarchitectures, which typically exhibit timing anomalies.

Calculation of Interference on Shared Resources. In the context of WCET analysis for multi-core processors, an important quantity is the amount of shared-resource interference that a concurrent processor core can generate while the program under WCET analysis is executed. To safely overapproximate worst-case interference generation scenarios, we generalized the well-known implicit path enumeration technique (IPET) [50] in a way that takes into account arbitrary subpaths of the abstract execution graph for the concurrent processor core [44]. As we assume that this generalized IPET does not scale to multiple real-world programs executed on a concurrent processor core, we sketched a program-modular calculation scheme [43] that calculates compositional base bounds per program on top of the generalized IPET.

Cache-Related Preemption Delay. In the presence of preemptive scheduling, preempting tasks evict cached memory blocks of preempted tasks, which have to be reloaded when the preempted tasks resume their execution [69]. This is commonly referred to as cache-related preemption delay (CRPD) [3]. We have experimentally evaluated the state-of-the-art techniques used to account for CRPD during timing analysis. Our experiments used task sets obtained by running LLVM on actual benchmarks. It turned out that the difference in precision of different CRPD analysis techniques and the overall impact of CRPD on schedulability are not as significant as observed for purely synthetically-generated task sets [69].

Strictly In-order Pipeline. To enable efficient and precise microarchitectural analysis, we introduced the strictly in-order pipeline in [35, 36]. This pipeline design enables an efficient progress-based abstraction and allows quantification of the effect of (amplifying) timing anomalies. The effect on WCET estimates and analysis performance have been evaluated using LLVM.

6 Conclusions

Since the kickoff of LLVM in 2014, it has been employed and extended in various research projects at Saarland University. It was exposed to TU Dortmund University in 2020. With the open source release of the code base in a version control system, and steady maintenance and integration, we foresee a great potential for LLVM to be used in research and education on WCET analysis more globally. Our short-term plan is to release practical exercises and tutorials for use of LLVM in education at the graduate level.

LLVM and a patched version of LLVM are available as open source for academic research purposes at:

https://gitlab.cs.uni-saarland.de/reineke/llvm
https://gitlab.cs.uni-saarland.de/reineke/llvm
References


2:14 LLVMTA: An LLVM-Based WCET Analysis Tool


