Validating the Reliability of WCET Estimates with MBPTA

Suzana Milutinovic[1,2], Jaume Abella[2], Francisco J. Cazorla[2,3]
[1]*Universitat Politècnica de Catalunya (UPC),*
[2]*Barcelona Supercomputing Center (BSC-CNS),*
[3]*Spanish National Research Council (IIIA-CSIC)*
suzana.milutinovic@bsc.es

Abstract-*Estimating the worst-case execution time (WCET) of tasks in a system is an important step in timing verification of critical real-time embedded systems. Measurement-Based Probabilistic Timing Analysis (MBPTA) is a novel and powerful method to compute WCET estimates based on measurements on the target platform. To provide reliable estimates, MBPTA needs to capture at analysis time the events with high impact on execution time. We propose a method to assess and increase the confidence that MBPTA captures the relevant events during analysis.*

I. Introduction

The worst-case execution time estimate (WCET) is an important metric in critical real-time systems to prove that each critical task will complete its function in time. The WCET estimates need to be reliable according to the domain-specific safety standard (e.g. ARP4761 in the avionics domain [1]), and as tight as possible to avoid wasting hardware resources during task scheduling. As real-time systems deploy increasingly complex hardware and software, satisfying both requirements for WCET estimation becomes a difficult challenge [2].

Measurement-Based Probabilistic Timing Analysis (MBPTA) [3] is a novel method to derive WCET estimates based on measurements on the target platform. MBPTA deploys Extreme value theory (EVT)[4], a statistical method used to describe tails of distributions. Based on a sample of collected measurements EVT returns the distribution of high execution times of a task with corresponding probabilities of exceeding them, Fig. 1. The pWCET estimate is the execution time value of that distribution that can only be exceeded with the cutoff probability selected in line with safety standard requirements.

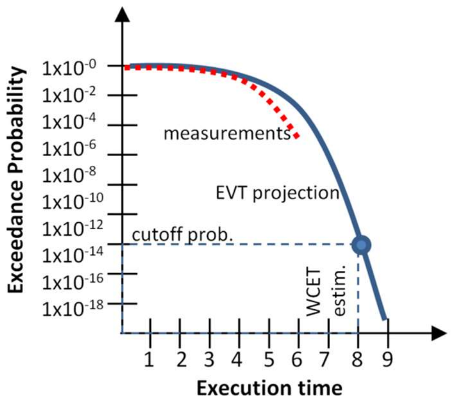

Fig. 1. Example of pWCET distribution.

MBPTA requires that the conditions under which the measurements during analysis are collected are equal or worse to the conditions at the system's deployment [5]. This requirement is ensured by: 1) forcing the sources that cause low variation of execution time to take its worst latency during analysis; and 2) time-randomizing the behavior of the sources that cause high variation of execution time. The main example of a hardware component whose timing behavior is randomized is a cache implementing random replacement and random placement policies.

The strength of MBPTA lies in the fact that it doesn't need to observe the longest execution time at a single measurement to predict that it can occur. If the worst outcome of each time-randomized resource is observed across different measurements, EVT will be able to predict the execution time when these bad scenarios occur together and upper-bound their probability of occurring simultaneously.

Our work focuses on deriving a method to guarantee with the specified level of confidence that the sample of measurements provided to EVT includes the events causing the worst outcomes of each time-randomized resource (we call them *events of interest*). As a consequence, the method guarantees that EVT will return a reliable pWCET estimate.

II. MBPTA: Capturing Events of Interest

In the previously studied architectures, MBPTA successfully captures the events of interest for each time-randomized resource, apart from time-randomized caches (TRc). TRc deploy the random placement policy, which maps addresses to randomly chosen sets. The mapping is changed across different runs, but kept constant during a single run. The authors in [6] observe that the execution time increases significantly if the number of addresses mapped to the same set exceeds the cache associativity. Some of these mappings, which we call *cache placements of interest*, may occur with a probability considered relevant by a safety standard, but low enough not to be captured during measurements at analysis time. Failing to provide the measurements capturing the cache placements of interest to the EVT method may cause the method to deliver optimistic pWCET estimates, and therefore return the unreliable results.

The HoG method proposed in [6] solves the problem of capturing cache placements of interest assuming that all addresses have the same impact on execution time. This is the case, for example, for a sequence that accesses all addresses in a round robin fashion. We extend such solution to the general case with arbitrary access patterns by proposing the Representativeness Validation by Simulation (RVS) method [7].

III. RVS METHOD

The RVS method analyzes the miss count impact f the different groups of addresses if they are placed in the same set by means of simulations and determines analytically the probability of this to occur. The validation is done in a miss count domain, as cache misses highly correlate with the execution times [7]. Then, by testing whether those pairs <miss count, probability> are upper-bounded by the pWCET curve obtained from applying MBPTA in the miss count domain with the default MBPTA number of runs (R), RVS can detect whether R runs are enough or, instead, extra runs are needed. Then RVS precisely identifies the number of runs needed (R’) so as to guarantee that all cache placements of interest have been observed

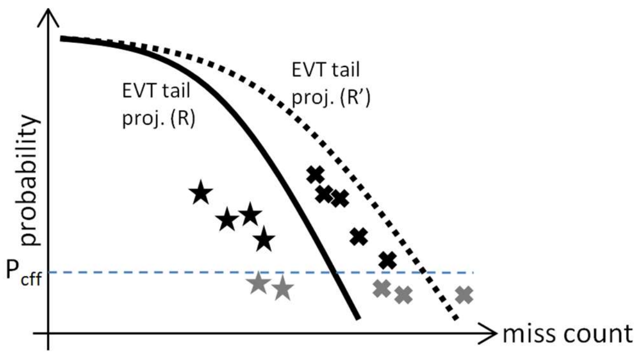

Fig. 2. Illustrative application of RVS.

with sufficient confidence.

This is illustrated in Fig. 2 where we show an example pWCET curve obtained with R runs that only upper-bounds some of the cache placements (marked with stars), but not some others (marked with crosses). Gray marks correspond to those cache placements that occur with negligible probability according to standards. RVS detects those cases leading to optimistic pWCET estimates and requests more runs (R’) so that they properly upper-bound all meaningful cache placements.

We show the result of applying the RVS method for *aifirf* benchmark from the EEMBC automotive suite [8], in Fig. 3. This particular benchmark fails to pass the validation step with R runs, but passes it successfully with R' runs, as determined by RVS, once the cache placements of interest are observed so that MBPTA can upper-bound them. In the figure we show the pWCET curve with R and R’ runs respectively, and the empirical complementary cumulative distribution function (ECCDF) for a large number of runs (orders of magnitude larger than R and R’).

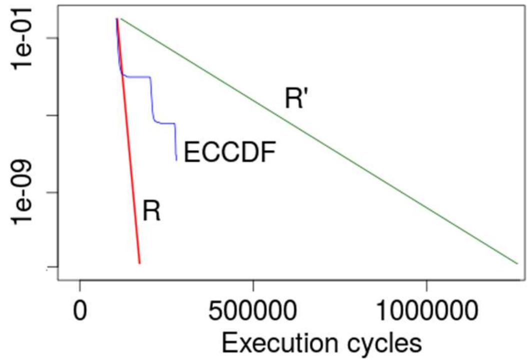

Fig. 3. pWCET for aifirf.

IV. CONCLUSIONS AND FUTURE WORK

MBPTA relies on EVT to estimate the pWCET of tasks, but to be reliable it requires that the execution time measurements used by EVT include all relevant (random) cache placements of interest. We propose the RVS method that determines the minimum number of measurements needed to feed MBPTA to produce reliable WCET estimates [7]. This is achieved by verifying that the pWCET estimate upper-bounds relevant cache placements in the miss domain. Our future work focuses on reducing the computation cost RVS and extending RVS toward multipath programs and cache hierarchies.

ACKNOWLEDGMENT

This work has received funding from the European Community’s FP7 programme [FP7/2007-2013] under grant agreement 611085 (PROXIMA, www.proxima-project.eu). Support was also provided by the Ministry of Science and Technology of Spain under contract TIN2015-65316-P and the HiPEAC Network of Excellence. Jaume Abella has been partially supported by the MINECO under Ramon y Cajal postdoctoral fellowship number RYC-2013-14717.

REFERENCES

[1] SAE International. ARP4761: Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment.

[2] J. Abella *et al*., "WCET analysis methods: Pitfalls and challenges on their trustworthiness," *10th IEEE International Symposium on Industrial Embedded Systems (SIES),* Siegen, 2015, pp. 1-10.

[3] L. Cucu-Grosjean *et al*., "Measurement-Based Probabilistic Timing Analysis for Multi-path Programs," *24th Euromicro Conference on Real-Time Systems (ECRTS),* Pisa, 2012, pp. 91-101.

[4] S. Kotz *et al*., "Extreme value distributions: theory and applications, " World Scientific, 2000.

[5] F.J. Cazorla *et al*. "Upper-bounding program execution time with extreme value theory, " In WCET Workshop, 2013.

[6] J. Abella *et al*. "Heart of Gold: Making the improbable happen to extend coverage in probabilistic timing analysis," In ECRTS, 2014.

[7] S. Milutinovic, J. Abella, F.J. Cazorla, "Modelling Probabilistic Cache Representativeness in the Presence of Arbitrary Access Patterns, ", In ISORC, 2016.

[8] Jason Poovey, "Characterization of the EEMBC Benchmark Suite," North Carolina State University, 2007.