

# Upper Bounds Prediction of the Execution Time of Programs Running on ARM Cortex-A Systems

Irina Fedotova , Bernd Krause, and Eduard Siemens

Department of Electrical, Mechanical and Industrial Engineering,  
Anhalt University of Applied Sciences, Bernburger Str. 57, 06366 Köthen, Germany  
{i.fedotova,e.siemens}@emw.hs-anhalt.de,  
bernd.krause@hs-anhalt.de

**Abstract.** This paper describes the application of statistical analysis of the timing behavior for a generic real-time task model. Using specific processor of ARM Cortex-A series and an empirical approach of time values retrieval, the algorithm to predict the upper bounds for the task of the time acquisition operation has been formulated. For the experimental verification of the algorithm, we have used the robust Measurement-Based Probabilistic Timing Analysis based on the Extreme value theory (EVT). In the ongoing work, we provide a systematic method for safe worst-case execution time estimations of programs running on the specific hardware. We focus here primary on the consequences of EVT assumptions and their correct interpretation for the upper bounds prediction.

**Keywords:** Extreme Value Theory · Worst-case execution time · Probabilistic Timing Analysis · Timing verification

## 1 Introduction

The implementation of most smart systems nowadays presumes the availability of high computing, memory and IO-resources on very small and low-power consuming hardware platforms - so called SoC devices. In this area, the ARM Cortex-A platform plays a central role. On such systems, the time acquisition process is one of the most critical steps in meeting real-time requirements. This presumes certain guarantees that the probability of the system failing to meet its timing constraints is below an acceptable threshold. However, the CPU of said system includes many features, which imposes correspondent limitations to the aimed timing analysis. One way to get suitable values for meeting real-time requirements is the development of statistical methods, predicting the probability distribution of the timing operations delays for particular tasks.

Nowadays, deterministic as well as probabilistic approaches pursue the goal of providing a safe upper bound of execution time for a particular task [1]. The difference between these approaches is mainly that the deterministic method produces a single worst-case execution time (WCET) estimation, while probabilistic - multiple WCET estimates with their respective probabilities. Both approaches have their static and measurement-based variants. Classical static timing analysis operates on deterministic

processor architectures and provide safe WCET estimates as they are proven to be the worst ones [2]. However, most real-time operating systems have become increasingly complex and it makes the applicability of static timing analysis is very challenging. On the contrary, measurement-based methods provide an estimation based on the derived maximal and minimal observed execution times or their distributions. In fact, with a probabilistic hardware architecture and measurement-based approaches, it is possible to guarantee an accurate probabilistic pWCET, what is solved by applying the Measurement-Based Probabilistic Timing Analysis (MBPTA).

The objective of the MBPTA is to derive the timing behavior of a task by means of statistical modeling. It aims at modeling extreme execution times for characterizing the worst-case, relied on both execution time measurements and the application of the Extreme Value Theory (EVT). The EVT deals with the extreme deviations from the median of probability distributions and estimates the tails, where the worst case should lie. However, hardware effects in real-time systems make EVT applicability difficult with regard to its required theoretical hypotheses. In this work, we focus on a particular hardware presented on the Atmel SAMA5D4 board with one ARM Cortex-A5 core, which can be a common example of realistic embedded applications. The platform based on these series of processors employs a cache of random-replacement policy. It facilitates upper bounding the probability of pathological worst-cases behaviors to extremely low levels (failure probability of  $10^{-9}$ ) [3]. For our investigations, the failure probability at the level  $10^{-9} < P < 10^{-7}$  (from hazardous class) was chosen. This level is proposed by PED certification [4] and usually applied in flight control systems.

In the related work, a number of methods for application of EVT theory for WCET estimations has been proposed [2, 5, 6]; nevertheless, a considerable uncertainty due to the complexity of the problem still exists. Results of statistical tests are often fuzzy and it is hard to make a correct decision on their basis to fulfill EVT requirements. To overcome difficulties with EVT, it is necessary to perform a complete set of diagnostic tests, check hypotheses for generalizing the EVT applicability and assume the validity of different methods. In this article, we intent to significantly reduce the existing ambiguities and suggest a systematic step-by-step method. Moreover, while previous studies have investigated time bounds for the whole application using corresponding benchmarks, we focus on the particular task for obtaining time values. The MBPTA approach is then applied only for this most critical part of the system, as other components can be assessed with more deterministic methods. The rest of the paper is organized as follows: the relation of the ongoing research to smart systems is discussed in Sect. 2, followed by discussion on related work in Sect. 3. The description of the problem of the real-time probabilistic modeling, as well as the main steps of proposed algorithm are presented in Sect. 4. Section 5 describes the experiment setup. The proof of fitting the target distribution and obtaining WCET estimators are provided in Sect. 6. Finally, Sect. 7 sums up and collects the conclusions and future work.

## 2 ARM Based SoC and Smart Systems

What principally differentiates the ARM Cortex-A series from others, is their wide use in performance-intensive systems. However, for our research the choice of these processors is based not only on the high performance and low power capabilities of the hardware, but also on the extensive application in control smart systems with timing constraints. The term “smartness” refers to the ability of the system performing very complex set of heuristic operations, based on a rich data basis, often gathered from external cloud environments within some given time boundaries, networking and real-time capabilities. Here selected embedded devices play a dominant role. A wide range of applications for Cortex-A is possible also due to the support of rich OSs such as Linux or Android. However, using them in embedded systems brings additional challenges to the development process. While one of the primary aim of any smart system is making decisions more predictive under certain conditions, in fact, the standard Linux OS is not productive enough for that. It follows therefore the necessity to better address some challenges related to the reaction of the system, especially where highly precise timing must be kept at one microsecond level.

The specific feature of these series of processor is, that CPU cycle counter is not directly available from user-space and wraps around in every 6.5 min. Thereupon, the investigations of timing capabilities are being performed within the high performance *HighPerTimer* library [7], which main idea is to simplify the process of timestamps acquisition. It allows avoiding the invocation of any system calls, since the main timing mechanism (supporting the procedure of handling overflows) is processed in user space. While microsecond precision of time stamping becomes standard for the development of remote monitoring and control systems, the underlying OS (particularly due to invoking system calls) is not able to meet the real-time requirements. Moreover, exactly the task for obtaining time values from the hardware can often take up to 20% of the whole program execution time. Therefore, it is important to minimize the cost of time retrieval and provide the precise time bounds of this critical task. Therefore, based on the used implementation of timing, it is possible to move forward in the direction of calculating worst-case execution time and predict the upper bound. These predictions allow a verification if the timing constraints will be met at run-time and so help to lead to a faster and more secure design process of the system.

## 3 Related Work

One of the first comprehensive comparisons among timing analysis techniques was presented by J. Abella et al. in [2]. Authors compare static and MBPTA methods in qualitative and quantitative terms under different cache configurations. Besides this, their previous work [8] establishes requirements to EVT with the MBPTA method to derive WCET estimates. Thereby they address the WCET problem by introducing randomization into the timing behavior of the system hardware and software.

The recent work by F. Guet et al. [6] proposes a DIAGnostic tool, which applies the MBPTA method without human intervention. Depending on the certain theoretical

hypotheses of the EVT, the logical work flow of the framework derives its probabilistic pWCET estimate from traces of execution times. Concerning the DIAGNOSTIC tool, K. Berezovskyi et al. in [9, 10] also investigate different methods inside the EVT theory for Graphical Processor Units (GPUs). The main results showed that hardware time-randomization is not that essential for the applicability of EVT and can be applied even to some non-time-randomized systems as GPUs.

An approach based on EVT theory has also been used for optimal performance analysis. Radojković et al. [5] presented a new method for predicting the performance of the thread assignment in multi-core processors. Using statistical inference of each thread assignment in a random sample, the authors estimates the optimal one.

Despite the intensive research in this area over the past few years, open problems in applying MBPTA for computing probabilistic WCET estimate still remain. In particular, said approaches give little information about the sequences of checking statistical hypotheses and making safe decision on their basis. Ongoing work intends to overcome these challenges by developing systematic and more coherent approach of analyzing the timing behavior. Moreover, the innovation novelty of this work is to cluster tasks exactly for time acquisition and apply the MBPTA method only for this critical non-deterministic part of the system.

## 4 The Probabilistic Modeling of Timing Behavior

Measurement-based methods produce estimates (for parameters of some distributions) by executing the given task on the given hardware and measuring the execution time of the task or of its parts. In particular, MBPTA approaches are aimed at modeling extreme execution times and characterizing its worst-case. The probabilistic theory that focuses on extreme values and large deviations from the average values is the EVT theory [11]. It estimates the probability of occurrence of extremely large values, which are known to be rare events. More precisely, the EVT predicts the distribution function for the max (or min) values of a set of  $n$  observations, which are modeled with random variables. Its main result is provided in the Fisher-Tippett-Gnedenko theorem, which characterizes the max-stable distribution functions for a single family of continuous cumulative distribution functions (CDFs), known as the generalized extreme value (GEV) distribution. The result implies that an average discrete cumulative distribution function belongs to the Maximum Domain of Attraction (MDA) of the GEV.

Furthermore, the GEV distribution is characterized by three parameters: location  $\mu$ , scale  $\sigma$ , and shape  $\xi$  and by their estimation, we can prove the resulting GEV distribution. If the shape parameter is  $\xi > 0$ , then the measured values in trace belong to Fréchet distribution, whereas if  $\xi < 0$ , then this is a “reversed” Weibull case. If the shape parameter  $\xi = 0$ , then values belong to a Gumbel distribution, which in most previous works [12, 13], has been assumed as applied to the pWCET distribution. Nonetheless, there is no restriction on the values that  $\xi$  can take and the resulting distribution, we intent to check all three distributions, in order to be close to the accurate estimation of the parameter. Therefore, for applying the EVT, the following hypotheses are required to be verified: that  $n$  random variables of execution time measurements are: (i) independent,

(ii) identically distributed and (iii) the existence of a distribution in the MDA of the GEV distribution is given [6]. These three elements are checked for guaranteeing the safety of the reliable probabilistic pWCET estimates.

#### 4.1 Selecting Extreme Values

Within the EVT context, there are several approaches to measure the extreme values. The Peak over Threshold (PoT) method is preferred and presents a good model for the upper tail, providing reliable extrapolation for exceedances over a sufficiently high threshold. The Pickands–Balkema–de Haan theorem [14], on which the PoT method is based, corresponds to the basic principle of extracting the execution time measurements from the trace above a threshold  $u$  and fitting of the Generalized Pareto Distribution (GPD) to the exceedances. The PoT allows to use data more efficient than other approaches, though the evident disadvantage is the selection of the suitable threshold value. Moreover, in the single-path case the PoT appears to be more accurate (with respect to the measurements), but the increase of the threshold  $u$  can result into more pessimistic pWCET estimations [14, 15].

#### 4.2 Proposed WCET Estimation Algorithm

Considering the probabilistic modelling briefly described above, the following algorithm for WCET estimation is suggested: **Step 1.** Selecting extreme values. The objective of this step is to collect from the original distribution the values, which fit into the tail, and hence can be modeled with the GEV distribution. For estimation of extremes, the PoT method was chosen: Step 1.1. The choice of the best-fitted threshold. It is based on the method of graphical diagnostic, which includes: mean residual life plot and parameter stability plot. They allow finding the lowest possible threshold, for which the extreme value model provides a reasonable fit to exceedances; Step 1.2. Filtering values above the threshold. **Step 2.** Fitting the GEV distribution: Gumbel, Fréchet and Weibull. If none of three types distributions fits, then going back to Step 1.1 and increase the threshold value. **Step 3.** Estimating the remaining parameters of the fitted distribution:  $\mu$ ,  $\sigma$  and  $\xi$ . **Step 4.** Verification of EVT hypotheses of independence and identical distribution: Step 4.1. Checking that the data are identically distributed; Step 4.2. Proving that samples are independent. **Step 5.** Return WCET estimation based on  $\mu$ ,  $\sigma$  and  $\xi$  parameter. Due to the space constrains, a more detailed description of the Step 4 is omitted here. However, the statistical analysis shows that EVT can be applied to the obtained results with high confidence (95%). At the same time, it must be considered that recent work [10] shows that independence is not a necessary hypothesis for EVT applicability and the theory can be applied even for stationary weakly dependent time series, even if the identical distribution of random variables does not represent a limiting hypothesis to EVT applicability.

## 5 Experimental Setup

There are multiple ways, in which measurements can be performed for the probabilistic approach. The most evident method is by extra instrumentation of the program code that collects a timestamp or CPU cycle counter. All experiments for this research have been carried out on ARM Cortex-A5 (see Sect. 2). We have evaluated the overhead of setting two consecutive timers of *HighPerTimer* library [7], which practically gives the estimate of the timer cost; the job, which is required to be performed in order to measure executions times of any program or the piece of code. The measured process is scheduled by `SCHED_OTHER` - the default scheduling policy of the Linux kernel version 4.4.11, used in these investigations. The influence of the real-time scheduling policy and kernel extensions are out of scope of this work. We rather concern about measuring the time acquisition process under the standard conditions. Table 1 lists basic statistics of the 27 min estimation part of a representative trace. The representation of complementary cumulative distribution function (CCDF) for this execution time is shown in Fig. 1(a). As can be seen from the graph, the distribution peaks near the mean and falls with rapidly decreasing probability density below 1  $\mu$ sec. The Fig. 1(b) gives the general picture of the time behavior showing additionally sigma values:  $\sigma$  and  $3\sigma$ .

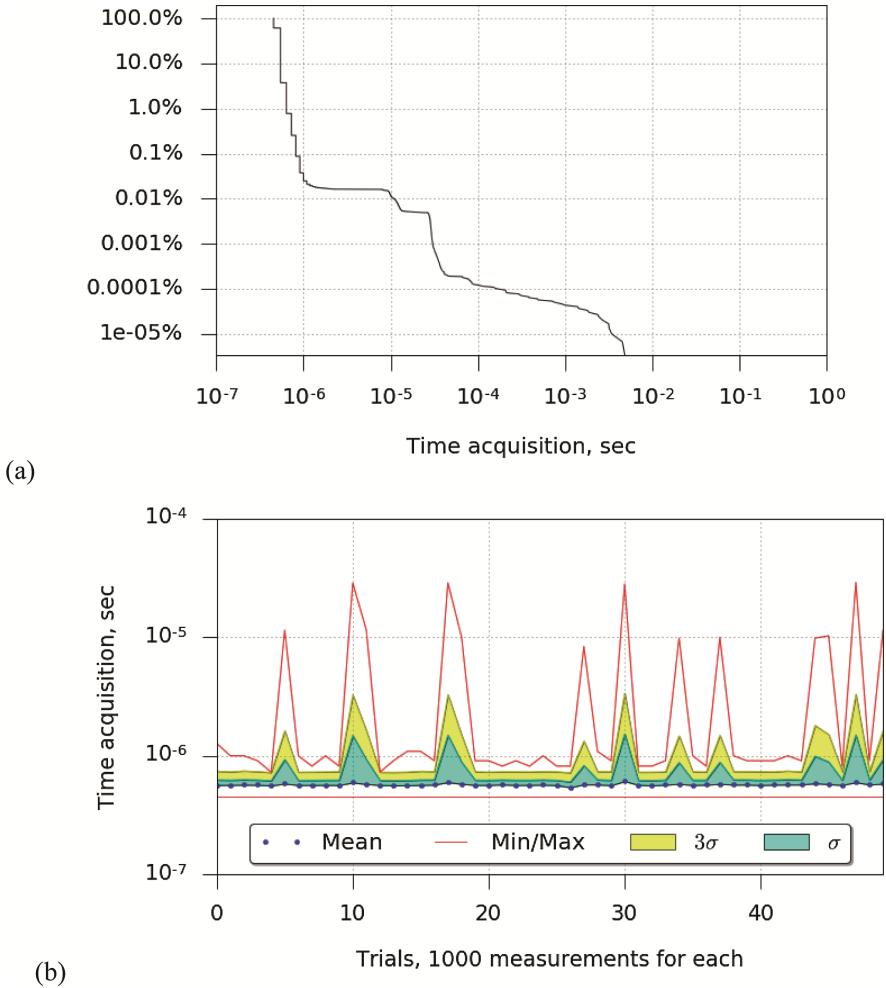
**Table 1.** Statistical properties of the original data set.

Mean execution time	Std. deviation	Max value	Min value
0.519 $\mu$ sec	2.217 $\mu$ sec	5.995 ms	0.455 $\mu$ sec

## 6 Fitting the GPD Distribution and Estimation the Upper Bounds

Having determined the threshold value, the parameters of the GPD can be calculated by the maximum likelihood estimation (MLE). The appropriate goodness-of-fit statistics compares observed data to quantiles of the specified distribution. We estimate three distributions by the critical value of Chi-square test, the Bayesian (BIC) and Akaike (AIC) information criteria and QQ-plot. Firstly, since preliminary test series show that the shape parameter  $\xi$  can't be negative, the case of Weibull distribution was excluded. Secondly, based on the Chi-square test the hypotheses that data follow Gumbel or Fréchet distributions can't be rejected. Thirdly, comparing models according to the AIC and BIC criteria, the hypothesis that data fit Fréchet distribution is slightly preferred against the Gumbel case. Fourthly, the straight diagonal line of a QQ-plot indicates that Gumbel distribution is a relatively good fit to the tail. However, since the difference of most statistics results for Gumbel and Fréchet models is not significantly differed, it makes sense to retrieve the WCET estimate for both cases.

The final step is to use the computed and verified GPD parameters and the exceedance probability of failure  $p$  to estimate the WCET. The WCET estimate is derived based on parameters  $\xi$  and  $\alpha_u$  for each respective scenario as following [16]:



**Fig. 1.** The original dataset of time execution: (a) its CCDF representation; (b) samples with statistical properties (mean, min/max,  $3\sigma$ ,  $\sigma$ ) grouped into 50 clusters.

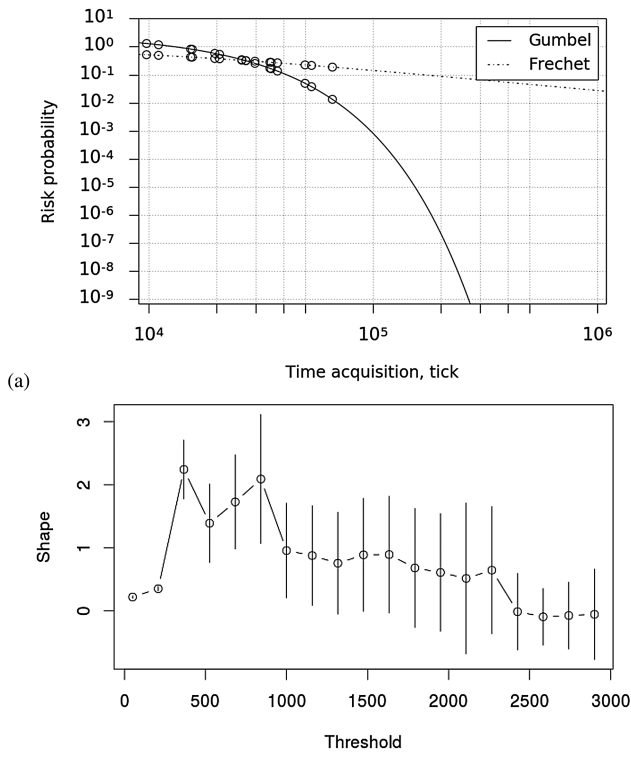
$$WCET = \begin{cases} u + \frac{\alpha_u}{\xi} \left( \left( \frac{n}{k} p \right)^{-\xi} - 1 \right) & \text{if } \xi > 0, \\ u - \alpha_u \log \left( \frac{n}{k} p \right) & \text{if } \xi = 0, \end{cases} \quad (1)$$

where  $\alpha_u$  is an estimated scale parameter,  $k$  - the number of peaks over the threshold  $u$ . Table 2 gives modeling results of the extreme execution times and Fig. 2 (a) shows the distribution convergence for Gumbel and Fréchet scenarios. The x-axis shows the pWCET estimation, the y-axis shows the respective probabilities. We use a CCDF of GPD to predict the potential values. According to the Table 2 the WCET estimates for both distributions differ significantly. Formally, too pessimistic results of Fréchet are

quite tight and practically less useful: for failure probability  $p = 10^{-9}$  the bound is 5.921 s. The Gumbel distribution converges to 0 faster than the Fréchet and it decreases the pessimism of the WCET thresholds. Moreover, shape parameter stability plot (on Fig. 2 (b)) estimates maximum likelihood of  $\xi$  over a range of threshold and confirms the hypothesis that data fit Gumbel distribution with  $\xi = 0$  respectively.

**Table 2.** EVT results for the hazardous class of  $p$  considering GPD expressed in cycles and msec.

	$\sigma$	$\mu$	$\xi$	(WCET; $10^{-7}$ )	(WCET; $10^{-8}$ )	(WCET; $10^{-9}$ )
Gumbel	12 247.56	12 968.35	0	27 968 = 2.54 ms	56 170 = 5.11 ms	84 370 = 7.67 ms
Fréchet	10 542.529	-1 539.657	1.351	124 185.2 = 11 ms	2 898 825 = 263 ms	65 126 941 = 5.921 s



**Fig. 2.** EVT results: (a) estimation of upper bounds; (b) shape parameter stability plot.

Having possibility to run the experiment including the target risk probabilities, we can compare the real max value and predicted WCET estimates. Therefore, the max value of the original dataset is 5.99 ms (Table 1) and the predicated upper bound using parameters for Gumbel distribution for failure probability  $p = 10^{-9}$  gives 7.67 ms, which fits the assumption. Considering any real-time system (soft or firm), these results allow to verify the timing constraints specifically for the application component, which



includes operations for time acquisition. This can help to design more stable and secure system running on platforms based on the ARM Cortex-A core.

## 7 Conclusion

The ability to predict the exceedance probability and reduce the cost to perform trustworthy timing analysis makes the MBPTA approach along with the EVT theory very attractive. The main contribution of this paper is to propose a systematics for the estimation of probabilistic WCET of a time acquisition task. Assuming several cases of GPD distributions, we compare WCET estimates for them. Our results show the least pessimistic estimation (about 7.67 ms for failure probability levels of  $10^{-9}$ ), using parameters for Gumbel distribution. Therefore, this method of predicting the upper bound is acceptable for the single-core ARM Cortex-A5 processor. For the future work, the investigation of other systems and hardware effects, such as the presence of multiple cores, cache memory effects or impacts of scheduling tasks by Linux kernel are planned. Considering hardware aspects should lead to significant improvements of the dependence metrics, reduction of the pessimism of the pWCET estimation and as a result, make applications, running on Linux, using ARM Cortex-A processors more time-predictable and so more suitable for being used in smart devices and applications.

## References

1. Wilhelm, R., Engblom, J., Ermedahl, A., Holsti, N., Thesing, S., Whalley, D., Bernat, G., Ferdinand, C., Heckman, R., Mitra, T.: The WCET problem - overview of methods and survey of tools. *ACM Trans. Embed. Comput. Syst.*, 36–53 (2007)
2. Abella, J., Hardy, D., Puaut, I., Quinones, E., Cazorla, F.J.: On the comparison of deterministic and probabilistic WCET estimation techniques. In: 26th Euromicro Conference on Real-Time Systems, Madrid, pp. 266–275 (2014)
3. Altmeyer, S., Cucu-Grosjean, L., Davis, R.I.: Static probabilistic timing analysis for real-time systems using random replacement caches. *Real-Time Syst.* **51**, 77–123 (2015)
4. Guidelines and methods for conducting the safety assessment process on civil airborne systems and equipment. ARP4761 (2001)
5. Radojković, P., Carpenter, P.M., Moretó, M., Čakarević, V., Verdú, J., Pajuelo, A., Cazorla, F.J., Nemirovsky, M., Valero, M.: Thread assignment in multicore/multithreaded processors: a statistical approach. *IEEE Trans. Comput.* **65**, 256–269 (2016)
6. Guet, F., Morio, J., Santinelli, L.: On the reliability of the probabilistic WCET estimates. In: 8th European Congress on Embedded Real-Time Software and Syst, Toulouse, pp. 758–767 (2016)
7. Fedotova, I., Siemens, E., Hu, H.: A high-precision time handling library. *J. Commun. Comput.* **10**, 1076–1086 (2013)
8. Cazorla, F., Vardanega, T., Quinones, E., Abella, J.: Upper-bounding program execution time with extreme value theory. In: the Proceedings of WCET workshop (2013)
9. Berezovskyi, K., Santinelli, L., Bletsas, K., Tovar, E.: WCET measurement-based and extreme value theory characterization of CUDA kernels. In: 22nd International Conference on Real-Time Networks and Systems, New York, pp. 279–288 (2014)

10. Berezovskyi, K., Guet, F., Santinelli, L., Bletsas, K., Tovar, E.: MBPTA is for graphics process or units. In 29th International Conference, Nuremberg, pp. 223–236 (2016)
11. Coles, S.: An Introduction to Statistical Modeling of Extreme Values. Springer, Heidelberg (2001)
12. Cucu-Grosjean, L., Santinelli, L., Houston, M., Lo, C., Vardanega, T., Kosmidis, L., Abella, J., Mezzetti, E., Quinones, E., Cazorla, J.F.: MBPTA for multi-path programs. In: 23rd Euromicro Conference on Real-Time Systems. IEEE Press (2012)
13. Hansen, J., Hissam, S.A., Moreno, G.A.: Statistical-Based WCET Estimation and Validation - [viewcontent.cgi](#). In 9th International Workshop on WCET Analysis (2009)
14. Balkema, A., Haan, L.: Residual life time at great age. *Ann. Probab.* **2**, 792–804 (1974)
15. Santinelli, L., Morio, J., Dufour, G., Jacquemart, D.: On the sustainability of the EVT for WCET estimation. In: 14th International Workshop on WCET Analysis, p. 21 (2014)
16. Embrechts, P., Klueppelberg, C., Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Springer Science & Business Media, Berlin (1996)