On the Sustainability of the Extreme Value Theory for WCET Estimation

Luca Santinelli[1], Jérôme Morio[1], Guillaume Dufour[1], and Damien Jacquemart[2]

1 ONERA – The France Aerospace Lab, Toulouse, name.surname@onera.fr
2 ONERA – The France Aerospace Lab, Palaiseau and INRIA Rennes ASPI

Abstract

Measurement-based approaches with extreme value worst-case estimations are beginning to be proficiently considered for timing analyses. In this paper, we intend to make more formal extreme value theory applicability to safe worst-case execution time estimations. We outline complexities and challenges behind extreme value theory assumptions and parameter tuning. Including the knowledge requirements, we are able to conclude about safety of the probabilistic worst-case execution estimations from the extreme value theory, and execution time measurements.

1998 ACM Subject Classification C.3 Special-Purpose and Application-Based Systems, C.4 Performance of Systems, G.3 Probability and Statistics

Keywords and phrases Extreme Value Theory, Worst-Case Execution Time, Probabilistic Worst-Case Execution Time, Dependence, Stationarity

Digital Object Identifier 10.4230/OASIcs.WCET.2014.21

1 Introduction

The measurement-based probabilistic timing analysis composes of measurements and statistic analyses, particularly the Extreme Value Theory (EVT), to outline rare events out of execution time measurements.

The EVT is a branch of statistics dealing with the extreme deviations from the median of probability distributions. It seeks to assess, from an ordered sample of a given random variable, the probability of events that are more extreme than any previously observed. The EVT applies to execution time and Worst-Case Execution Time (WCET), since it offers guarantees to have projected the tail of measurement distributions. Thus, EVT provides what is called safe WCET estimations accounting for rare events.

Real-time relies on WCETs to model task execution behaviors: a real-time system becomes predictable by always accounting for the worst-case at every task execution. As the input space for the task code is finite and the hardware behavior is assumed to be deterministic, it is reasonable to argue about the exact worst-case execution time. The WCET_{exact} and its estimation $\overline{C}$ are upper-bounds for any possible execution behavior of that task.

Unfortunately, systems are unpredictable as the environment can be diverse and dynamic with multiple possible evolutions in time. Both hardware and software elements may experience some variability or even randomness[1], e.g. multi-processor, cache, branch predictors, DRAM refresh, interruptions that occur whenever they are most inappropriate, and the interferences between interacting elements in the system lead to the dependences

[1] Randomness is intended in the common sense, as lack of pattern or predictability in events.

14th International Workshop on Worst-Case Execution Time Analysis (WCET 2014).
Editor: Heiko Falk; pp. 21–30
OpenAccess Series in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

that emphasize unpredictabilities and variability, [7]. WCET_{exact} is in general unknown and potentially unknowable, and the estimation $\overline{C}$ could be extremely pessimistic.

Probabilities could model system indeterminacy and unpredictabilities, capturing multiple behaviors with their frequencies of happening. Such a fine grained system probabilistic representation reduces the pessimism brought by deterministic models, where only the estimation of the worst-case is taken into account. Then, the challenge is to guarantee probabilistic estimations of worst-case execution time, and build with the probabilities a safe alternative to the deterministic real-time. The EVT has been recently applied with that purpose, but it is still far from accomplishing such job.

1.1 State of the Art

A measurement-based approach to timing analysis implies a near-zero cost for system coding and modeling in order to get task execution time observations [2]. The problem consists of guaranteeing the coverage of all the possible execution conditions and obtaining reliable WCET estimations.

Ensuring exhaustive execution condition coverage requires knowledge of the system, slightly reducing the advantage of measurement-based approaches with respect to static timing analysis approaches [14]. Even if the worst-case execution time is computed using test generation techniques which ensure feasible path coverage, usual assumptions reduce the input set variability. This fact decreases complexity, but demands an improved system model to identify the worst-case execution condition.

More "analytical" approaches to measurement-based timing analysis, make use of the statistics of extremes [8, 9] to construct predicted WCETs. In particular, recent works have re-formalized the application of EVT to the WCET problem [2, 1] by making use of an ad-hoc probabilistic hardware architecture. Those approaches are able to guarantee accurate probabilistic WCET estimations from measured execution time distributions.

Contributions. In this article, we intend to show challenges and possibilities applying the EVT to the task execution problem. Thus, we present the EVT resulting distribution for different system conditions and parameters. Besides, we aim at continuing the discussion around claimed robustness of the extreme value theory in terms of the guarantees it offers to the probabilistic worst-case execution time estimation.

The idea of this work is to present some results from a set of experiments to show: i) the impact of EVT parameters on the resulting worst-case execution time distributions; ii) the differences between block maxima EVT and peak over threshold EVT, the two approaches to EVT; iii) a qualitative evaluation of the EVT robustness. We focus on the required EVT hypotheses and their impact on the resulting pWCET estimations. The whole statistical analysis framework is to begin a complete and formal discussion about EVT sustainability to the execution time problem.

Data Setup. For our tests we make use of real traces taken from an Intel(R) Xeon(R) E5620 2.4 GHz dual socket, each socket with four cores and three levels of cache. The schedMcore[2] runtime support and Linux Trace Toolkit new generation (LTTNG) tracing framework are applied to guarantee real-time execution and extract accurate execution time measurements. The task implementations considered are single-path and multi-path tasks from the Mälardelen benchmark suite[3]. Out of them we extract the execution time measurements as traces called "single-path" and "multi-path", respectively.

[2] `https://forge.onera.fr/projects/schedmcore`
[3] `http://www.mrtc.mdh.se/projects/wcet`

Furthermore, we apply two artificial execution time random distributions as the normal distribution and a multi-modal distribution (obtained from 3 normal random variables combined). They are considered to compare results with realistic measurements. All the execution time are in *nsec*, and the number of observations is 100000 per trace.

2 Probabilistic Timing Analysis

Execution Time Profiles (ETPs) $\mathcal{C}$ are measured execution time distributions.[4] $\mathcal{C}$ are empirical distributions that lack of completion and coverage, which means that from the ETPs it is not possible to conclude about worst-case execution time. Nonetheless, ETPs are important to investigate in order to derive execution pattern/trends from which define worst-case execution conditions, and then worst-case execution times.

2.1 Probabilistic Worst-Case Execution Time

Statistical estimations of the worst-case execution time induce the notion of probabilistic WCET (pWCET), alternative to the deterministic WCET. pWCET is as distributions of WCET values C_j, each of them with associated a probability p_j of being WCET, $\mathcal{C} = \begin{pmatrix} C_j \\ p_j = P\{\mathcal{C} = C_j\} \end{pmatrix}_{j \in \{1,\cdots,J\}}$; p_j is the probability that C_j is an upper-bounds of the task execution time. The following is a possible definition of pWCET.

▶ **Definition 1** (probabilistic Worst-Case Execution Time)**.** Given $\mathcal{C}_k$ the distribution of execution time measured in a certain configuration/condition k. The probabilistic Worst-Case Execution Time distribution $\mathcal{C}^*$ of a task is a tight upper bound on the execution time $\mathcal{C}_k$ of all possible execution conditions. Hence, $\forall k,\ \mathcal{C}^* \succeq \mathcal{C}_k$.[5]

The exact pWCET $\overline{\mathcal{C}}^*$ would be the tightest upper bound to any $\mathcal{C}_k$, $\mathcal{C}^* \succeq \overline{\mathcal{C}}^* \succeq \mathcal{C}_k,\ \forall k$.

The probabilistic worst-case execution time can also be defined from the exceeding thresholds and the 1-Cumulative Distribution Function (1-CDF) representation. Given a probability of exceedence p^*, the value C^* is the worst-case execution time such that $P\{\mathcal{C}^* \geq C^*\} \leq p^*$. Alternative to the pWCET distribution, we can call minimum probabilistic worst-case execution time the tuple $\langle C^*, p^* \rangle$. In real-time and certification issues, safety is validated with exceedence probability smaller than 10^{-9}. We consider that threshold and define the pWCET as $\langle C^*, 10^{-9} \rangle$, although the following reasoning is open to any threshold.

A pWCET estimation $\mathcal{C}^*$, in order to be safe, has to be greater than or equal to the exact pWCET, which is unknown, and any measurement $\mathcal{C}_k$. A distribution $\mathcal{C}^*$ is greater than or equal to a distribution $\mathcal{C}_k$ iff $P\{\mathcal{C}^* \leq c\} \leq P\{\mathcal{C}_k \leq c\}$ for any c and the two random variables are not identically distributed (two different distributions). For the $\langle C^*, 10^{-9} \rangle$ estimation, we say that it is safe if for each pWCET $\langle C, p \rangle$, including the exact one, for all $C \geq C^*$, $p \leq 10^{-9}$.

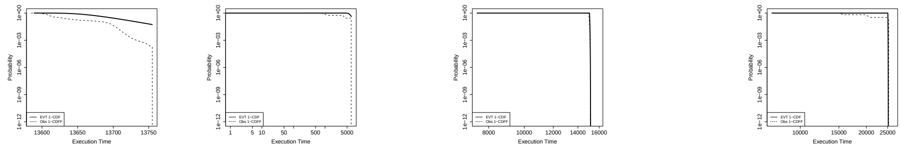
2.2 Extreme Value Theory

The extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions. It seeks assessing, from a given ordered sample of

---

[4] We use calligraphic letters to represent probability distributions; non calligraphic letters are for single values.

[5] Total ordering of distributions is guaranteed by comparing distribution probabilities. Thus a distribution $\mathcal{C}_i$ is greater than or equal to a distribution $\mathcal{C}_j$, $\mathcal{C}_i \succeq \mathcal{C}_j$, iff $P\{\mathcal{C}_i \leq c\} \leq P\{\mathcal{C}_j \leq c\}$, for every c.



Single-path-BM Multi-path-BM Artifact single-path-PoT Artifact multi-path-PoT

Figure 1 EVT applied to different cases.

a random variable, the probability of events that are more extreme than any previously observed. The EVT allows to estimate tails of distributions and thus explores rare events, where the WCET and its probabilistic version pWCET should lie. Two are the EVT approaches possible: Block Maxima (BM) and Peak over Threshold (PoT). The safety of the statistical pWCET estimation through EVT has always been referred to the independence and identical distribution (i.i.d.) hypotheses. If both are verified, then the EVT distribution tail projection can be considered as a safe pWCET estimation. Along this paper we discuss those assumptions and extend them.

Figure 1 shows the tail projection effect of the EVT once applied to measured distributions. With the 1-CDF representation we appreciate not strong difference between the measurements distributions and the EVT distributions. For the scope of the paper, this is what we call accuracy of the EVT estimations.

2.2.1 Law of sample maxima – Block maxima approach

EVT is notably very useful when one has to work with only a fixed set of data (measurements/observations), not having other info outside those observed. Consequently it is assumed in the following that a set of i.i.d. samples $X_1, \dots, X_N$ of a time series $(X_t)_{t>0}$ (equivalently as distribution $\mathcal{X}$ being the samples from distributions[6]) is available. The associated ordered sample set is defined with $X_{(1)}, \dots, X_{(N)}$. EVT enables to estimate for some thresholds S the probability $P\{(X_t)_{t>0} > S\}$.

The main result of EVT [5], is that the maxima of an i.i.d. sequence converges to a Generalized Extreme Value (GEV) distribution G_ξ under some general conditions, which admits a generic Cumulative Distribution Function (CDF) $\mathcal{G}_\xi(x)$. GEV distributions are composed of three distinct types, characterized by $\xi = 0$, $\xi > 0$ and $\xi < 0$ that correspond to the Gumbel, Fréchet and Weibull distributions respectively. Let us define $\mathcal{G}$, the CDF of the i.i.d. samples $X_{(1)}, \dots, X_{(N)}$.

▶ **Theorem 2.** *Suppose there exist a_N and b_N, with $a_N > 0$ such that, for all $y \in \mathbb{R}$ $P\left\{\frac{\mathcal{X}_{(N)}-b_N}{a_N} \leq y\right\} = \mathcal{G}^N(a_N y + b_N) \xrightarrow{N\to\infty} \mathcal{G}(y)$, where $\mathcal{G}$ is a non degenerate CDF, then $\mathcal{G}$ is a GEV distribution $\mathcal{G}_\xi$. In this case, one denotes $\mathcal{G} \in MDA(\xi)$ (MDA = Maximum Domain of Attraction).*

Unless samples of maxima are directly available, it is then required to group the samples $X_{(1)}, \dots, X_{(N)}$ into blocks and fit the GEV using the maximum of each block. While the a_N and b_N distribution parameters are found by best fitting the input trace of events, the grouping into block maximum is somewhat an arbitrary parameter. Although all three

[6] A sample, equivalently an observation, comes from a distribution, thus the representation as calligraphic or non-calligraphic letters are equivalent.

parameters are influent, we focus on the block size, named b from now on, to show the impact that it has on the EVT pWCET estimation.

2.2.2 Peak over threshold approach

Instead of grouping the samples into block maxima, PoT considers the largest samples X_i to estimate the probability $P\{\mathcal{X} > S\}$. The link between EVT and the distribution of a threshold exceedence is firstly described in [13]. The following theorem can be obtained:

▶ **Theorem 3.** *Let us assume that the distribution function $\mathcal{G}$ of i.i.d. samples $X_1, \ldots, X_N$ is continuous. Set $y^* = \sup\{y, \mathcal{G}(y) < 1\} = \inf\{y, \mathcal{G}(y) = 1\}$. Then, the next two assertions are equivalent: a) $\mathcal{G} \in MDA(\xi)$, and b) there exists a positive and measurable function $u \mapsto \beta(u)$ such that $\lim_{u \mapsto y^*} \sup_{0<y<y^*-u} |\mathcal{G}^u(y) - \mathcal{H}_{\xi,\beta(u)}(y)| = 0$. $\mathcal{G}^u(y) = P\{\mathcal{X} - u \leq y | \mathcal{X} > u\}$, and $\mathcal{H}_{\xi,\beta(u)}$ is the CDF of a generalized Pareto distribution (GPD) with shape parameter ξ and scale parameter $\beta(u)$.*

The expression of the GPD distribution function $\mathcal{H}_{\xi,\beta}(x)$ depends on both ξ and β.

Theorem 3 is in fact useful to estimate a probability of exceedance $P(\mathcal{X} > S)$ since it can be rewritten as $P\{\mathcal{X} > S\} = P\{\mathcal{X} > S | \mathcal{X} > u\} \cdot P\{\mathcal{X} > u\}$, for $S > u$. A natural estimate of $P\{\mathcal{X} > u\}$ is given by the empirical mean $\hat{P}^{MC}\{\mathcal{X} > u\} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{X_i > u}$, from the Monte Carlo method (MC). With the Theorem 3 and for significant value of u, one obtains $\hat{P}\{\mathcal{X} > S | \mathcal{X} > u\} = 1 - \mathcal{H}_{\xi,\beta(u)}(S - u)$. The estimate of $P\{\mathcal{X} > S\}$ is then built with $\hat{P}^{POT}\{\mathcal{X} > S\} = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{X_i > u}\right) \cdot \left(1 - \mathcal{H}_{\xi,\beta(u)}(S - u)\right)$.

The parameter u can be selected arbitrarily, as it affects the accuracy as well as safety of the EVT PoT estimation.

3 Extreme Value Theory applicability

First we describe what is the independence we are looking for in order to apply the EVT, then we extend it to other conditions. The meaning of independence we are looking for is whether individual observations within the same execution trace are correlated with each other or not. Knowing one observation tells you something about another, in which case they are dependent; knowing one observation tells you nothing about another, in which case they are independent. In this section we investigate deeply stationarity and independences/dependences for execution time traces.

The independence is not a necessary hypothesis for the EVT, since Leadbetter et al. [11] and Hsing [10] developed EVT for stationary weakly dependent time series. Those two references also establish statistical tools for that situation. For independent but not identically distributed random variables, a basic probabilistic result is in Mejzler [12]. In our EVT investigation, we start considering and supporting less constraining hypotheses than independence, such as stationarity and extremal dependences. We present tests to verify them and the guarantees that can be provided to the results of the extreme value theory with such assumptions.

3.1 Stationarity

Given execution time data sets, to verify stationarity the autocorrelation can be computed with lag plots, or turning point test can be performed. These are to model the relationship that exists between measured observations.

Table 1 Independence, stationarity and identical distribution tests.

	runs test	Kolmogorov-Smirnov test	autoregressive	Ljung-Box test
single-path	0.506	0.02804	AR(1)	0.07153
multi-path	0.6832	0.5434	AR(7)	0.2169
normal mono-variate	0.5755	0.8222	AR(0)	0.4593
multi-variate	0.729	0.4653	AR(0)	0.9825

Hypothesis testing means deciding, from a number of observations, whether one should consider a property to be true or not. The resulting ρ-value tells whether accept or reject the null hypothesis H_0. Normally, $\rho > 0.05$ validates the null H_0, while $\rho \leq 0.05$ rejects H_0, thus validates the alternative H_1.

With no mean of formalism, a process is stationary if its mean, variance and autocovariance structure do not change over time. An autoregressive (AR) model is a representation of type for random process: i) $AR(0)$ denotes the sequence of observations without dependence – white noise, ii) $AR(1)$ is process with a positive φ parameter where only the previous observation in the process and the noise term contribute to the output, and so on. We make also use of the Ljung-Box (LB) test, which examines whether there is significant evidence for non-zero correlations between lags. Large ρ-values from the LB test suggest that the series is not stationary, thus there is no trend between consecutive observations; this would support independence.

A test applied in [2] aims at proving that samples are independent by looking for randomness. This is called runs test, or Wald Wolfowitz test, where randomness is sought within the observed data series by examining the frequency of "runs"; a "run" is a series of similar responses. Furthermore, we consider the Kolmogorov-Smirnov (KS) test to verify the identical distribution hypothesis and check if the observations follow the same distribution.

Table 1 describes the hypothesis verification tests and their results with respect to the input measurements. For runs test, KS test and LB test, results are given as ρ-value. Noticeably, real execution traces are independent (single-path) or stationary (multi-path). The EVT can be applied to real cases, since their randomicity is enough for the EVT application. This means that there would not be the need for extra randomicity, as for example with random replacement cache policies, [2, 1].

3.2 Dependence of the extreme samples

While showing that just stationarity is needed to apply EVT, we can still get independence as far as extreme samples are concerned. Indeed, EVT can still be applied on time series with temporal dependence if the extreme samples are sufficiently separated in time. In that case, extreme samples can be considered as independent. To estimate the dependence level of extreme samples, it can be interesting to compute the extremogram of the samples.

An extremogram is a measure of extremal dependence for time series measurement [3]. Contrary to the usual methods for characterizing the dependence of samples, it only focuses on their extreme values. Let us firstly consider the theoretical definition of an extremogram.

The extremogram $\rho(h)$ of a stationary time series $(X_t)_{t\geq 0}$ is defined by: $\rho(h) = \lim_{n\to+\infty} \frac{P\{\mathcal{X}_0 > a_n, \mathcal{X}_h > a_n\}}{P\{\mathcal{X} > a_n\}}$, with a_n a sequence such that $P\{|\mathcal{X}| > a_n\} \approx n^{-1}$. The variable h can be seen as a correlation length.

a) Single-path time series b) Multi-path time series

Figure 2 Extremogram of single-path time series and multi-path time series examples.

In practice, if one assumes that the consecutive samples $(X_t)_{1 \leq t \leq N}$ are available, an estimator $\hat{\rho}(h)$ of the extremogram $\rho(h)$ has been proposed as $\hat{\rho}(h) = \frac{\sum_{t=1}^{N-h} \mathbf{1}_{(X_t > a, X_{t+h} > a)}}{\sum_{t=1}^{N} X_t > a}$, where $\mathbf{1}_{(X_t > a, X_{t+h} > a)}$ is equal to 1 if $(X_t > a, X_{t+h} > a)$ and 0 otherwise. The threshold a is experimentally set as the 0.98-quantile of the samples.

Several remarks can be made on the estimator $\hat{\rho}(h)$. Firstly, this estimator varies between 0 and 1. When $\hat{\rho}(h) \longrightarrow 1$, the extremal samples are highly correlated. In that case, group of consecutive extremal samples can be observed in time. De-clustering algorithm is then often required in order to apply safely EVT [6]. When $\hat{\rho}(h) \longrightarrow 0$, the extremal samples are uncorrelated and can arise as individual sample in the time series. EVT can then be applied with more confidence.

Figure 2 shows the extremogram to single- and multi-path cases. The extreme samples of these 2 time series have a limited correlation, $\hat{\rho}(h) < 0.1$, and de-clustering is not required.

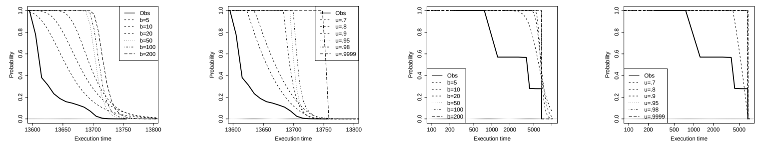
4 Extreme Value Theory approaches

The idea behind the EVT is close to a black box approach which is applied to avoid knowledge of the system, thus overcoming the complexity that today's systems have. The advantage that EVT offers to pWCET estimation is about the relatively small cost and good accuracy of the pWCET estimations. Unfortunately there exists complexity due to parameter selection. Indeed, in both BM and PoT cases, there are parameters to be defined (respectively b and u), and their impact to the resulting pWCET has to be considered. In this section we depict the differences that exist between the two forms of EVT applied to task execution time.

4.1 EVT parameters

To evaluate the impact of block size and threshold parameters we apply EVT by changing those parameters within a certain range. For BM, the size of the blocks b is such that $b \in \{5, 10, 20, 50, 100, 200\}$. For PoT, the thresholds are selected via the quantiles $q(p)$, where p is the quantile probability. Thus, it is $u \in q(0.7), q(0.8), q(0.9), q(0.95), q(0.98), q(0.9999)$. Figure 3 compares the parameters effects with four different input distributions and the 1-CDF representations. For the empiric distributions there is more accuracy from the EVT estimations, at least within a certain exceeding probability range $[1, 10^{-6}]$. While in the single-path case the PoT appears to be more accurate[7], in case of multi-path execution traces, it is BM which is more accurate. The effectiveness of the EVT depends on the shape of the input distribution, and it is not possible to conclude about one EVT approach being better than another. To note how increasing the parameters, i.e. increasing the block size or the threshold (the quantile), the quality of the EVT estimation degrades not linearly.

[7] Accuracy is empirically defined with respect to the measurements, since the exact pWCET is unknown.



BM single-path PoT single-path BM multi-path PoT multi-path

Figure 3 EVT pWCET distributions and values varying block size and threshold.

Table 2 10^{-9} exceeding thresholds with different EVT parameters compared with the measured value.

	max	BM $b = 5$	BM $b = 10$	BM $b = 20$	BM $b = 50$	BM $b = 100$	BM $b = 200$
single-path	13755	14388	14312	14184	13886	13894	13918
multi-path	6750	41812	26353	19065	6779	6820	6920
	max	PoT $p(0.7)$	PoT $p(0.8)$	PoT $p(0.9)$	PoT $p(0.95)$	PoT $p(0.98)$	PoT $p(0.9999)$
single-path	13755	13758	13758	13769	14006	14729	13752
multi-path	6750	6861	6861	6974	34563	$4.94 * 10^8$	6851
	max	$u, p(0.7)$	$u, p(0.8)$	$u, p(0.9)$	$u, p(0.95)$	$u, p(0.98)$	$u, p(0.9999)$
single-path	13755	13623	13636	13684	13695	13701	13751
multi-path	6750	3985	6697	6712	6714	6715	6832.33

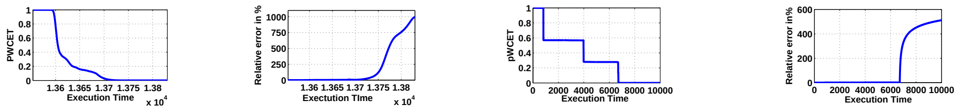
In Table 2 it considered the pWCET tuple definition $\langle C^*, 10^{-9}\rangle$. Here is understandable the difficulty that EVT has in accurately estimating the pWCET with extreme parameters, such as small b and PoT with few values above the threshold ($u = 0.9$ or more). Estimated C^* at limit cases are either very far from the measured values, i.e. multi-path with $u = 0.95$, $u = 0.98$, and $b \leq 20$, or very close to the measured values, i.e. $b = 100$ and $b = 200$. Both PoT and BM pWCET estimation remains safe with no underestimations. But while increasing b the pessimism decreases, and perhaps the capability of embedding rare events decreases, increasing u it is the accuracy of the pWCET estimation to reduce resulting into more pessimistic pWCET estimations. In case of multi-path those trends are more evident due to the multi-variate distribution the EVT has to handle. In there, interesting are the results for $u = 0.98$ and $u = 0.9999$, respectively with a huge pWCET estimation and a maybe too small one. Critical cases have to be avoided in order to PoT safe and sound. Last three rows of the table are the thresholds used by the PoT at the respective quantiles, to give an idea about where are the estimated pWCETs.

Both table and figure describe outline the complexity in selecting the best parameter, which depends on the input measured distribution.

4.2 EVT robustness: estimation of relative error

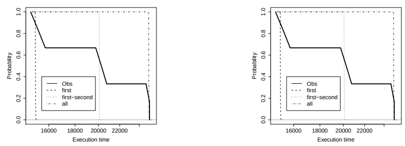
Bootstrap [4] is a well-known statistical method that enables to estimate characteristics of a statistics. It can notably be applied to estimate relative error and confidence interval of the pWCET distribution and quantiles obtained with EVT.

For that purpose, it is firstly needed to re-sample the data $X_1, \ldots, X_N$ to obtain a bootstrap re-sample $X_1^*, \ldots, X_M^*$. The term X_j^* is a sample set of size N and is determined from $X_1, \ldots, X_N$ by random sampling with replacement. From the bootstrap re-sample $X_1^*, \ldots, X_M^*$, one can then estimate M pWCET distributions, $\mathrm{pWCET}_1, \ldots, \mathrm{pWCET}_N$ and the associated 10^{-9}-quantiles obtained with EVT. The relative errors or the 95% confidence interval of these quantities are then easily computable.



single-path dist. sigle-path relative error multi-path dist. multi-path relative error in %

Figure 4 pWCET and associated relative of single-path and multi-path time series examples.



BM and multi-path PoT and multi-path

Figure 5 EVT path variation and frequency impact.

Let us apply bootstrap to estimate relative errors and confidence interval for single- and multi-path series with $M = 100$. The pWCET distribution and its corresponding relative error are plotted in Figure 4 for single-path series. The 10^{-9} quantile of the pWCET distribution is equal to 13766 with $[13743, 13797]$ as 95% confidence interval. The pWCET distribution and its corresponding relative error are plotted in Figure 4 for multi-path series. The 10^{-9} quantile of the pWCET distribution is equal to 6922 with $[6730, 7454]$ as 95% confidence interval. In both single- and multi-path series, the pWCET distribution obtained with bootstrap is smooth since it is estimated with a mean operator. The accuracy of the pWCET distribution decreases with execution time, indeed the probability associated to pWCET becomes very low when execution time increases. This probability is thus badly estimated since the number of samples N is constant.

4.3 EVT robustness: completeness

In this section we show some of the limits of the EVT once applied to the worst-case execution time problem. It is due to the knowledge of the system: in order to be effective (and safe), the EVT has to know which are the worst case execution conditions.

In Figure 5 we have applied the artificial multi-modal distribution as input to the EVT. The multi-modal distribution has been obtained combining three normal distributions with different mean values. In our test we have changed the execution time inputs. “first” is the measurement trace obtained from the normal distribution with the smallest mean. The EVT applied to that trace of observations is labeled “first”. “first-second” is the case where the EVT is applied to a trace obtained with the least mean and second least mean normal distributions. “all” has the whole set of observations from the three distributions. We notice that, just with partial information (not the whole multi-modal distribution but portions of it), the EVT is not able to safely infer the extremes, and thus pWCETs.

The EVT needs complete set of inputs (depicting the measurement conditions) in order to be safe: EVT robustness depends on the knowledge of the system and its execution conditions, including variability sources due to input variability and path coverage.

5 Conclusion

With this work we begin a formal analysis of the extreme value theory applied to the probabilistic worst-case execution time problem. We first present the problem as well as the EVT and its two possible approaches, i.e. block maxima and peak over threshold. Their applicability together with assumptions are verified, proving that the i.i.d. hypothesis is too strict, while stationarity and extremal dependences are allowed for safe EVT pWCET estimations. Furthermore, we provide initial verification means to EVT complexity and parameter selection, outlining the impact that parameters have on the pWCET estimation. Finally, we introduce the notion of robustness for EVT estimations.

In the future, we intend to continue in those directions aiming at listing the EVT limits and its potential. This helping the developer better choosing between measurement-base approach and static timing analysis, and perhaps combining both in an efficient hybrid timing analysis.

Acknowledgement. The work of Damien Jacquemart is financially supported by DGA (Direction Générale de l'Armement) and Onera.

References

1 Francisco J. Cazorla, Tullio Vardanega, Eduardo Quinones, and Jaume Abella. Upper-bounding program execution time with extreme value theory. In *WCET*, 2013.
2 L. Cucu-Grosjean, L. Santinelli, M. Houston, C. Lo, T. Vardanega, L. Kosmidis, J. Abella, E. Mezzeti, E. Quinones, and F. J. Cazorla. Measurement-Based Probabilistic Timing Analysis for Multi-path Programs. In *23nd Euromicro Conference on Real-Time Systems (ECRTS)*. IEEE, 2012.
3 R. A. Davis and T. Mikosch. The extremogram: A correlogram for extreme events. *Bernoulli*, 2009.
4 B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY, 1993.
5 P. Embrechts, C. Kluppelberg, and T. Mikosch. Modelling extremal events for insurance and finance. *ZOR Zeitschrift for Operations Research Mathematical Methods of Operations Research*, 97(1):1–34, 1994.
6 Christopher A. T. Ferro and Johan Segers. Automatic declustering of extreme values via an estimators. *EURANDOM report, Eindhoven University of Technology*, 2002.
7 Mark K. Gardner. *Probabilistic analysis and scheduling of critical soft real-time systems*. PhD thesis, University of Illinois at Urbana-Champaign, Champaign, IL, USA, 1999. AAI9953022.
8 E. J. Gumbel. *Statistics of Extremes*. Columbia University Press, 1958.
9 J. Hansen, S. Hissam, and G. A. Moreno. Statistical-based wcet estimation and validation. In *the 9th International Workshop on Worst-Case Execution Time (WCET) Analysis*, 2009.
10 T. Hsing. On tail index estimation using dependent data. *The Annals of Statistics*, 1991.
11 M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 1983.
12 D. Mejzler. On the problem of the limit distribution for the maximal term of a variational series. *Lvov. Politehn. Inst. Naucn Zap. Ser. Fiz.-Mat.*, 1956.
13 J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3(1):119–131, 1975.
14 R. Wilhelm, J. Engblom, A. Ermedahl, N. Holsti, S. Thesing, D. B. Whalley, G. Bernat, C. Ferdinand, R. Heckmann, T. Mitra, T. Mueller, I. Puaut, P. P. Puschner, J. Staschulat, and P. Stenström. The worst-case execution-time problem – overview of methods and survey of tools. *ACM Trans. Embedded Comput. Syst.*, 2008.