

Challenging the Fallacy of Exponential Tail Upper-Bounding in Probabilistic Real-Time Systems.

Sandeep Kumar Mishra

Department of Electronics and Telecommunication, College of Engineering Bhubaneswar

Abstract— As a probabilistic real-time computing method, measurement-based probabilistic timing analysis is based on extreme value theory (EVT), a statistical theory that is applied to worst-case execution time (WCET) analysis on real-time embedded systems. The statistical distribution produced by the EVT theory is known as the generalized Pareto distribution or the generalized extreme-value distribution. One of the three conceivable shapes for their cumulative distribution function asymptotically could be: 1) light; 2) exponential; or 3) hefty tail. A number of recent papers suggested using their exponential version to upper-bound the light-tail distributions. We demonstrate in this letter that this assumption is frequently misunderstood and that it is only true in specific situations. Because of this, the WCET is estimated in an unsafe manner, which is unacceptable for applications that target safety-critical embedded systems.

Index Terms—Embedded software, real-time systems, statistical distributions.

I. INTRODUCTION

PROBABILISTIC real-time computing has been proposed to overcome the complexity of using traditional worst-case execution time (WCET) analyses on modern processor architectures. In fact, advanced micro-architectural features—such as pipelines, multilevel caches, or system management interrupts—complicate the WCET estimation [2], being sources of nondeterminism. The timing analyses become even more challenging on commercial off-the-shelf platforms [3] and in time-sensitive high performance computing applications [1]. In critical embedded systems, the violation of timing constraints is not acceptable. To guarantee their satisfaction, it is mandatory to get a *safe*, i.e., nonunderestimated, value for the tasks WCET.

The widely used probabilistic approach is called measurement-based probabilistic timing analyses (MBPTAs).

This method has still several open issues [4] and it has recently gained considerable research interest. It is based

on using direct measurements of the task execution times

i.e., the probability of observing an execution time that is larger than the maximum value previously observed. The output of probabilistic real-time analysis is the so-called probabilistic-WCET (pWCET), i.e., a statistical distribution used to derive the WCET, given the violation probability, and vice versa. Some works upper-bound this probability with the exponential version of its distribution. In this letter, we argue that this assumption is valid only under certain conditions. A comprehensive and basic tutorial on probabilistic real-time can be found in [5].

A. Extreme Value Theory

The extreme value theory (EVT) [6] is a well-known and assessed statistical theory, traditionally used for natural disaster prediction, e.g., the forecast of catastrophic floods. Provided that the input values are i.i.d., the output of the EVT is a distribution that represents the probability of observing extreme minimums or maximums.

In real-time computing, the EVT input values, identified by the random variables X_1, X_2, \dots, X_n , are the time measurements of a given task's jobs. Following a well-defined estimation and testing process [7], the EVT can estimate the pWCET cumulative distribution function (cdf): $F(x) = P(X < x)$. Its complementary (ccdf), i.e., $1 - F(x) \geq P(X > x)$, represents the probability of experiencing execution times larger than a fixed value x .

B. Extreme Distributions

From the results of the Fisher-Tippett-Gnedenko theorem [8], [9], the pWCET distribution is asymptotically equivalent to the generalized extreme value distribution (GEV or GEVD)

$$F(x) = \begin{cases} e^{-\left(\frac{x-\mu}{\sigma}\right)^{\xi}} & \text{if } \xi \neq 0 \\ e^{-e^{-\frac{x-\mu}{\sigma}}} & \text{if } \xi = 0 \end{cases}$$

to predict the occurrence of extreme events at runtime,

The GEV has three parameters: 1) the location μ ; 2) the scale σ ; and 3) the shape ξ . The value of the shape parameters specializes the distribution: for $\xi < 0$, the distribution is a reversed Weibull; for $\xi = 0$, it is a Gumbel; for $\xi > 0$, it is a Fréchet. Traditionally, the estimation of the GEV par-

ameters is performed by using the block-maxima (BM) method: given the original sequence of random variables X_1, X_2, \dots, X_n and a block size B , BM generates the

¹Independent and identically distributed. This assumption can be relaxed, but this discussion is outside the scope of this letter.



Fig. 1. Asymptotic behavior of the GP3(0, 100, ξ) cdf $F(x)$.

following sequence $Y_1, Y_2, \dots, Y_{n/B}$ of random variables:

$$Y_i = \max(X_{B \cdot (i-1) + 1}, X_{B \cdot (i-1) + 2}, \dots, X_{B \cdot (i-1) + B}).$$

The peak-over-threshold (PoT) approach is an alternative to BM

$$(Y_1, Y_2, \dots, Y_m) = \{X_{s.t. X > u}\}$$

where u is a predefined threshold and m is the size of the right term set. The resulting random variables Y_i are distributed according to the generalized Pareto distribution (GPD) in the PoT case. The GPD can be provided in three-parameter form (GP3) or two-parameter form (GP2). In order to simplify the subsequent notations, we consider the GP2 lacking the location parameter μ , i.e., GP2(σ, ξ) GP3(0, σ, ξ). This form does not reduce the generality of this letter, as discussed in Section II-C. The cdf of GP2(σ, ξ) is therefore defined as

$$F(x) = \begin{cases} 1 - 1 + \xi x^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - e^{-x/\sigma} & \text{if } \xi = 0. \end{cases} \quad (1)$$

The GPD is asymptotically equivalent to GEV [10] and it can consequently be used for the extreme probability computation.

II. EXPONENTIAL TAIL-BOUNDING PROBLEM

In probabilistic real-time research, Hansen *et al.* [11], Cucu-Grosjean *et al.* [12], and Abella *et al.* [13] argued that the exponential-tail distributions, i.e., GEV/GPD with $\xi = 0$,

in this context, this means that using a distribution with exponential-tail ($\xi = 0$) to upper-bound a light-tail ($\xi < 0$) should not lead to pWCET underestimation. This is true because, for a certain WCET x , the probability to incur in a longer execution time ($1 - F_{G_1}(x)$) is always higher than the one computed with light-tail ($1 - F_{G_2}(x)$). Vice versa, for a given probability p , the WCET estimated by exponential tail ($F_{G_1}^{-1}(1 - p)$) is always higher than the one computed by light-tail ($F_{G_2}^{-1}(1 - p)$).

²These results are valid only if the other parameter(s) of the distribution remains unchanged. For example, assume to fit the complete distribution GP3(μ, σ, ξ) (with $\xi < 0$) and then enforce $\xi = 0$ obtaining GP3($\mu, \sigma, 0$). In this case, the second distribution upper-bounds the first one and the pWCET is not underestimated. However, enforcing $\xi = 0$ before performing the distribution fitting may lead to different values of μ and σ , with respect to the real distribution of the data. This would invalidate the previous result, carrying out potentially unsafe pWCET estimations. The estimation procedure in fact usually provides the μ and σ values that best fit the input data. These in general are different from the ones that would have been computed without enforcing $\xi = 0$. Moreover, if the estimator is unbiased with respect to the mean of the extreme population, the resulting distribution is always unsafe, as proven below.

A. Parameters Shift Effects on the First Moment

To prove the statements above, we initially use at two parameters GPD GP2(σ, ξ) (this to simplify the calculus). The extension onto GP3 and GEV is then discussed in Section II-C.

C. Let Y_1, \dots, Y_m be the m maximum time measurements distributed under $G \sim \text{GP2}(\sigma, \xi)$, as a result of the EVT, e.g. by using the PoT algorithm, $Y_i > u$. If the estimator is unbiased, the mean value of Y_i matches the expected value of the GP2 distribution

$$E[Y] = \frac{\sigma}{1 - \xi}.$$

It follows that, for the exponential distribution case ($\xi = 0$)

the expected value is:

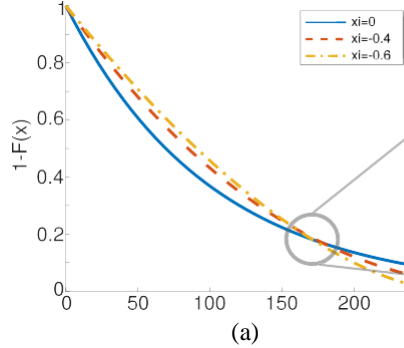
$$E[Y] = \sigma_{\xi=0}.$$

are good candidates for fitting the WCET distribution. These claims are motivated by empirical demonstrations, as it is almost impossible to provide formal ones. Conversely, other experiments [14], [15] showed that it is worth considering the $\xi < 0$ and $\xi > 0$ cases. The exponential distribution may in fact not be representative of all the scenarios.

In literature, the $\xi > 0$ case is controversial: considering it would mean that the WCET can also get an infinite value. However, discussing the validity of this case is out of the scope of this letter.

Fig. 1 shows the tails of the extreme distributions, depicted by the cdf for different values of the shape parameter ξ . This result has been recently used [16]–[18] to state that exponential-tail distribution upper-bounds the light-tail distribution. Formally,

$1 - F_G(x) > 1 - F_G(x)$
Upper-bounding with the exponential-tail distribution means forcing $\xi = 0$ for the same set of data, while maintaining the same expected value $E[Y]$, that is,



$\xi = 0 \quad 1 - \xi$
In case the data are distributed with $\xi < 0$, the simplification $\xi = 0$ leads to estimate $\sigma_{\xi=0} > \sigma_{\xi < 0}$. However, the following section provides the proof that this scale parameter skew may lead to unsafe WCET estimations.

B. Proof of Failure of Exponential-Tail Upper-Bounding

Given the definition of cdf $F(x) = P(X \leq x)$, upper bounding a distribution in the context of MBPTA means that the relation $F^r(x) < F(x)$ holds for any x .

Proof: Let $\bar{F} = 1 - F(x) = P(X > x)$ be the complementary where $\xi = 0 \sim GP3(\mu, \sigma, 0)$ and $\xi < 0 \sim GP3(\mu, \sigma, \xi < 0)$. This relation is equivalent for GEV. In the probabilistic real-time tardy cdf. A safe upper-bound for WCET has to guarantee the conservative relation $P^r(X > x) > P(X > x) \forall x$. From this,

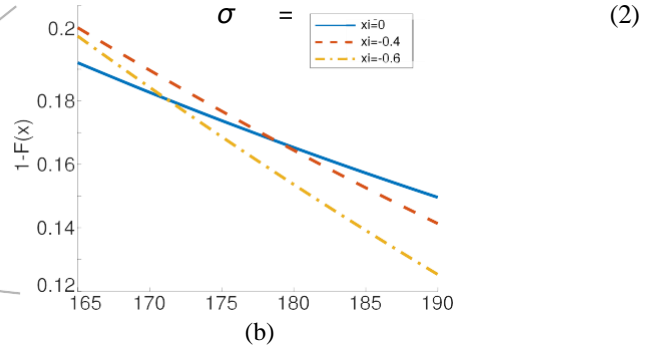


Fig. 2. Complementary cdf varying ξ of two-parameter GPD. The referenced distribution GPD(100, 0) is represented by the solid blue line. The (b) plot is the zoom of (a) plot at the cdf intersection.

it is possible to obtain $F^r(x) > 1 - F(x)$ and, in turn, $F(x) < F^r(x)$.

In our scenario, $F^r(x)$ corresponds to the upper-bound with $\xi = 0$, while $F(x)$ is the real distribution, with ξ assumed to be set to an unknown negative value $\xi < 0$. By expanding the cdfs we obtain

$$1 - e^{-\frac{x}{\sigma}} \quad \xi = 0 < 1 - \frac{x}{1 + \xi} \frac{x^{-1/\xi}}{\sigma_{\xi < 0}} \quad (3)$$

This inequality must hold for any x , but since we are dealing with a positive variable (execution time), this holds only for $x > 0$. As a consequence of (2), it is possible to state that this inequality is not true in general.

Proof: Removing the constant term and multiplying by -1

$$e^{-\frac{x}{\sigma}} \quad \xi = 0 > 1 + \xi \frac{x^{-1/\xi}}{\sigma_{\xi < 0}} \quad (4)$$

C. Applicability to GEV and GPD3-Parameters

The problem of exponential-tail bounding exists also in GP3 and GEV since both share the same tail behavior presented in Fig. 1. The previous proofs can be easily ported to GP3 and GEV distributions. Following the same approach of the GP2 version, the mean value of $GP3(\mu, \sigma, \xi)$ is $E[Y] = \mu + (\sigma / (1 - \xi))$. If the mean value does not change once the estimation run with $\xi = 0$, the results are exactly the same of the provided proof. If $\mu_{\xi=0} < \mu_{\xi < 0}$, then the error is higher and

the estimation becomes unsafe. If $\mu_{\xi=0} > \mu_{\xi < 0}$, then nothing can be said without an accurate analysis of the specific case.

Similarly to the GPD case, the GEV condition for safe upper-bounding is

$$e^{1 - e^{-\frac{x}{\sigma_{\xi=0}}}} < e^{-\frac{x}{\sigma_{\xi < 0}}}$$

and since $e^{f(x)} < e^{g(x)} \Leftrightarrow f(x) < g(x)$, it results that

$$-\frac{x}{\sigma_{\xi=0}} < -\frac{x}{\sigma_{\xi < 0}}$$

Let us now replace $\sigma_{\xi < 0}$ according to (2) $\sigma_{\xi < 0}$

$$\frac{x}{\sigma_{\xi < 0}} > 1 + \frac{\xi x^{-1/\xi}}{\xi \sigma_{\xi < 0}} \quad (5)$$

The equation corresponding to this inequality has a trivial zero for $x_1 = 0$, but it has another solution \bar{x}_2 for $x > 0$

$$\bar{x}_2 = \frac{\sigma_{\xi < 0}}{\xi} - W(\xi - 1) e^{-\frac{1}{\xi} - 1} + \xi - 1 \quad (6)$$

where $W[\cdot]$ is the Lambert W function. Since $\xi < 0$, the argument of $W[\cdot]$ is negative, as well as $W[\cdot]$. While, being $W[\cdot] < 1$, \bar{x}_2 assumes a positive value. This means that there is a second zero ($x_2 > 0$) and consequently at least a value (x_2) that violates the inequality of (3).

As discussed in Section III, there is actually a continuous

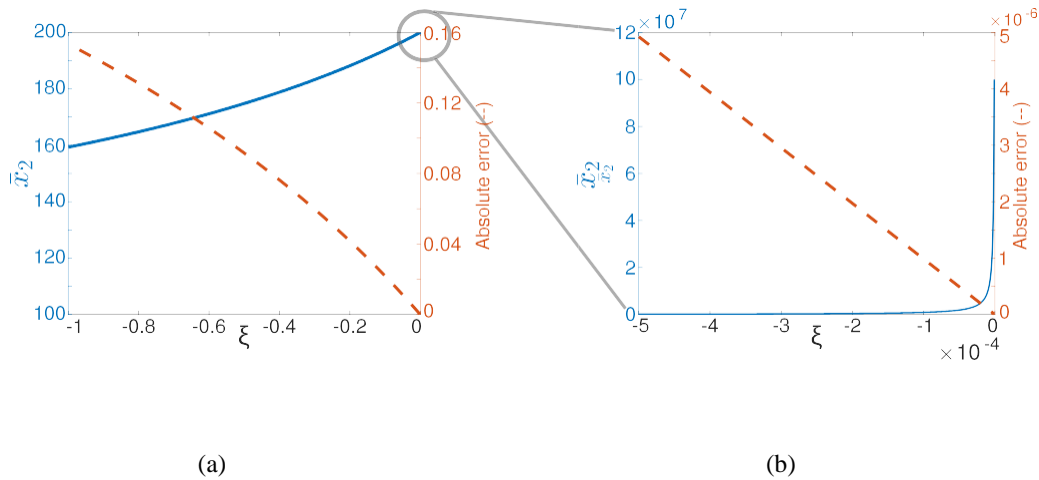


Fig. 3. Analysis of the safety point x_2 (blue solid line) and the maximum absolute error (red dashed line), varying thereal $\xi < 0$ and compared to $\xi = 0$ case. The (b) plot is the zoom of (a) plot when ξ has a magnitude of 10^{-4} .

and $\xi = 0.8$. The scale parameter σ is computed according to (2). The respective ccdfs are depicted in Fig. 2(a). The exponential GPD ($\xi < 0$) upper-bounds both distributions only starting from the value $x_2 = 179$. The absolute value of x_2 is not negligible: in this case we consider $\sigma = 100$, that is the mean value of the extremes, and the exponential tail upper-bound becomes safe only after nearly the double of it. Fig. 2(b) zooms in the intersection point of the previous Fig. 2(a). As expected, increasing the value of ξ toward 0 produces a smaller error in the difference of cdf between $\xi < 0$ and $\xi = 0$. However, it shifts also the intersection point, i.e., the point at which the upper-bound is safe, toward $+\infty$.

To investigate better the last result, we compute the maximum error between a $GP2(\sigma_{\xi < 0}, \xi < 0)$ and the

$$1 - e^{-\xi x} \quad \xi < 0 < 1 - 1 + \xi$$

that is exactly the same of (3). For this reason, the previous analysis can be applied also to GEV.

III. NUMERICAL EVALUATION

In order to clarify the previous equations and to provide a counterexample to the exponential upper-bounding claim, we use the GP2 distribution with $\sigma = 100$ and $\xi = 0$ as a reference. To show the violation of the upper-bounding rule, we compare it with the two GP2 distributions with $\xi = -0.4$ interval, i.e., infinite points, that violates the inequality. Providing an analytical proof for it is not simple because of the complexity introduced by the Lambert W function. Luckily, this is not necessary to demonstrate that the exponential-tail upper-bounding is unsafe. The counterexample obtained by numerical evaluation is in fact sufficient to prove this.

referenced distribution $GP2(100, 0)$, by varying the value of ξ from 1 to 0. The result is depicted in Fig. 3(a). The blue (solid) line represents the value x_2 after which the upper-bound is safe, while the red (dashed) line shows the maximum error compared to the reference distribution. The x_2 value increases with a peculiar slope, for which we provide Fig. 3(b) to show the trend for small values of ξ , while the maximum error has a quasi-linear trend. The key point here is to observe that there is no upper-bound for x_2 . Even if there exists a point x_2 from which the upper-bounding is safe, it is not possible to know it without knowing the real value of ξ . This leads to an uncertainty on the pWCET estimation that cannot be accepted in hard real-time systems. On the other hand, when ξ is close to 0 and x_2 increases toward infinite, the error decreases, but we still need to know ξ in order to estimate both.

IV. CONCLUSION

To ensure timing limits in harsh real-time embedded systems, the WCET analysis's dependability is crucial. Given that it provides pWCET estimates, probabilistic real-time computing presents a viable approach to handling complicated infrastructures. A few recent studies have suggested an upper bound for the pWCET extreme value distribution when it has an exponential tail version ($\xi = 0$) and a light tail ($\xi < 0$). Although it streamlines the procedure in general, if the ξ value is bounded a priori with regard to the estimation phase, the pWCET value can be underestimated. Because they assumed the upper boundary to be safe, previous works ignored this issue. Rather, we demonstrated that its validity is limited to WCET values higher than an unidentified value $\times 2$. These arguments make it inappropriate for us to discuss exponential-tail distributions on crucial systems without taking the parameter into consideration. Instead, either the error size needs to be predicted or the WCET needs to be sufficiently big to ensure a safe upper bound.

REFERENCES

- [1] W. Fornaciari, G. Agosta, D. Atienza, E. A. C. Brandolese, *et al.*, "Reliable power and time-constraints-aware predictive management of heterogeneous exascale systems," in *Proc. Int. Conf. Embedded Comput. Syst. Archit. Model. Simulat. (SAMOS)*, Jul. 2018.
- [2] C. Berg, J. Engblom, and R. Wilhelm, "Requirements for and design of a processor with predictable timing," in *Proc. Perspectives Workshop Design Syst. Predictable Behav. Dagstuhl Seminar (IBFI)*, 2004, p. 1–20.
- [3] D. Dasari, B. Akesson, V. Nélis, M. A. Awan, and S. M. Petters, "Identifying the sources of unpredictability in core-based multicore systems," in *Proc. 8th IEEE Int. Symp. Ind. Embedded Syst. (SI ES)*, Jun. 2013, pp. 39–48.
- [4] S. J. Gil *et al.*, "Open challenges for probabilistic measurement-based worst-case execution time," *IEEE Embedded Syst. Lett.*, vol. 9, no. 3, pp. 69–72, Sep. 2017.
- [5] F. Reghenzani, G. Massari, and W. Fornaciari, "The extreme value theory in probabilistic real-time computing," Politecnico di Milano, Milan, Italy, Oct. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.1462957>
- [6] L. De Haan and A. Ferreira, *Extreme Value Theory: An Introduction*. New York, NY, USA: Springer, 2007.
- [7] F. Reghenzani, G. Massari, and W. Fornaciari, "Chronovise: Measurement-based probabilistic timing analysis framework," *J. Open Source Softw.*, vol. 3, no. 28, p. 711, 2018. [Online]. Available: <https://doi.org/10.21105/joss.00711>
- [8] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Math. Proc. Cambridge Philos. Soc.*, vol. 24, no. 2, pp. 180–190, 1928.
- [9] B. Gnedenko, "Sur la distribution limite du terme maximum d'une série aleatoire," *Ann. Math.*, vol. 44, no. 3, pp. 423–453, 1943.
- [10] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values*. Basel, Switzerland: Birkhäuser, 2007.
- [11] J. Hansen, S. Hissam, and G. A. Moreno, "Statistical-based WCET estimation and validation," in *Proc. 9th Int. Workshop Worst Case Execution Time Anal.*, vol. 10, 2009, pp. 1–11.
- [12] L. Cucu-Grosjean *et al.*, "Measurement-based probabilistic timing analysis for multi-path programs," in *Proc. 24th Euromicro Conf. Real Time Syst.*, Jul. 2012, pp. 91–101.
- [13] J. Abella, E. Quiñones, F. Wartel, T. Vardanega, and F. J. Cazorla, "Heart of gold: Making the improbable happen to increase confidence in MBPTA," in *Proc. 26th Euromicro Conf. Real Time Syst.*, Jul. 2014, pp. 255–265.
- [14] L. Santinelli, F. Guet, and J. Morio, "Revising measurement-based probabilistic timing analysis," in *Proc. IEEE Real Time Embedded Technol. Appl. Symp. (RTAS)*, Apr. 2017, pp. 199–208.
- [15] G. Lima, D. Dias, and E. Barros, "Extreme value theory for estimating task execution time bounds: A careful look," in *Proc. 28th Euromicro Conf. Real Time Syst. (ECRTS)*, Jul. 2016, pp. 200–211.
- [16] J. Abella, M. Padilla, J. D. Castillo, and F. J. Cazorla, "Measurement-based worst-case execution time estimation using the coefficient of variation," *ACM Trans. Design Autom. Electron. Syst.*, vol. 22, no. 4, p. 72, Jun. 2017.
- [17] L. F. Arcaro, K. P. Silva, and R. S. D. Oliveira, "On the reliability and tightness of GP and exponential models for probabilistic WCET estimation," *ACM Trans. Design Autom. Electron. Syst.*, vol. 23, no. 3, p. 39, Mar. 2018.
- [18] I. Agirre *et al.*, "Fitting software execution-time exceedance into a residual random fault in ISO-26262," *IEEE Trans. Rel.*, vol. 67, no. 3, pp. 1314–1327, Sep. 2018.