Probabilistic Real-Time Systems

Liliana Cucu-Grosjean

Abstract—Proposed originally as stochastic scheduling, the probabilistic real-time scheduling concerns real-time systems with at least one parameter described by a random variable. Any parameter of the task may have such description, but the existing work concentrates on the probabilistic worst-case execution times. This paper will provide the main results for such systems in the case of one processor and a list of the open problems. The second part of the lecture deals with real-time systems that have stochastic description of the parameters. All the probabilistic operations are based on convolutions and the complexity of these operations may be a problem for realistic implementations. Sampling techniques providing decreased complexity are presented. The lecture ends with the presentation of the main open directions for probabilistic real-time systems and their impact on real-time systems in general.

Index Terms—Real-time, probabilistic, worst-case.

I. INTRODUCTION

The arrival of more complex and modern architectures for critical embedded systems imposes the use of analyses overcoming the requirement to have a certain knowledge of the worst-case execution time (WCET) of a task. The probabilistic² approach is one of the approaches overcoming this requirement. Since the seminal paper of Lehoczky [1], probabilistic reasoning has been the basis for response time analysis [2], worst-case execution time estimation [3] or for the proposition of optimal priority assignment policies [4]. Probabilistic approaches take into account the fact that extremely large values for parameters like worst-case execution time or minimal inter-arrival time are rare. In this paper we discuss results for worst-case execution times. The idea that the worst-case execution time of a task may have a low probability of occurrence is first presented by Burns and Edgar [5] and it is later confirmed by results [6], [7], [8] heavily inspired by this first work [5].

In addition to the proposition of a new technique for estimating the WCET of tasks, the work of Burns and Edgar proposes the original definition of what is currently known as probabilistic worst-case execution time (pWCET) estimate (a formal definition is provided in Section II-B). This notion is different from the probabilistic execution times (pETs) [9] and the pWCET may be used as input to the probabilistic schedulability analyses like [2] [10] [11] [12] [13]. This difference between pWCET and pET has an important consequence on the independence between the random variables describing the WCET as described in [14].

In this paper we list the main probabilistic real-time analyses in the case of one processor and a list of the open problems.

This paper is organized as follows. In Section II we introduce the notations, the definition of pWCET and pET and their relation. The three notions of independence are detailed and discussed in Section III. In Section IV we present the related work with respect to this paper and we conclude in Section V.

II. NOTATIONS

We consider a task set of \( n \) tasks \( \{\tau_1, \tau_2, \ldots, \tau_n\} \). Each task \( \tau_i \) is characterized by four parameters \((O_i, C_i, T_i, D_i)\) where \( O_i \) is the arrival of \( \tau_i \), \( T_i \) is the minimal inter-arrival time (commonly known as period) and \( D_i \) the relative deadline. The variable \( C_i \) is a random variable² and it describes the probabilistic worst-case execution time (pWCET) of task \( \tau_i \). We denote by \( C_i^j \) the probabilistic execution time (pET) of the \( j^{th} \) job of \( \tau_i \).

A random variable \( \mathcal{X} \) has associated a probability function (PF) \( f_{\mathcal{X}}(\cdot) \) with \( f_{\mathcal{X}}(x) = P(\mathcal{X} = x) \). The possible values \( X^0, X^1, \ldots, X^k \) of \( \mathcal{X} \) belong to the interval \([x_{\text{min}}, x_{\text{max}}]\), where \( k \) is the number of possible values of \( \mathcal{X} \). In this paper we associate the probabilities to the possible values of a random variable \( \mathcal{X} \) by using the following notation

\[
\mathcal{X} = \left( \begin{array}{cccc}
X^0 = X^{\text{min}} & X^1 & \cdots & X^k = X^{\text{max}} \\
f_{\mathcal{X}}(X^0) & f_{\mathcal{X}}(X^1) & \cdots & f_{\mathcal{X}}(X^{\text{max}})
\end{array} \right),
\]

where \( \sum_{j=0}^{k} f_{\mathcal{X}}(X^j) = 1 \). A random variable may be also specified using its cumulative distribution function (CDF) \( F_{\mathcal{X}}(x) = \sum_{x=x_{\text{min}}}^{x} f_{\mathcal{X}}(z) \).

A. Probabilistic Execution Time

Definition 1. The probabilistic execution time (pET) of the job of a task describes the probability that the execution time of the job is equal to a given value.

For instance the \( j^{th} \) job of a task \( \tau_i \) may have a pET

\[
C_i^j = \begin{pmatrix}
2 & 3 & 5 & 6 & 105 \\
0.7 & 0.2 & 0.05 & 0.04 & 0.01
\end{pmatrix}
\]

(2)

If \( f_{C_i}(2) = 0.7 \), then the execution time of the \( j^{th} \) job of \( \tau_i \) has a probability of 0.7 to be equal to 2.

B. Probabilistic Worst-case Execution Time

Definition 2. The probabilistic worst-case execution time (pWCET) of a task describes the probability that the worst-case execution time of that task does not exceed a given value.

For instance, a task \( \tau_i \) may have a pWCET

²In this paper we use a calligraphic typeface to denote random variables, e.g., \( \mathcal{X}, \mathcal{C}, \) etc.
\[
C_i = \begin{pmatrix}
2 & 3 & 105 \\
0.8 & 0.19 & 0.01
\end{pmatrix}
\] (3)

If \( f_C(2) = 0.8 \), then the worst-case execution time of \( \tau_i \) has a probability of 0.2 to be larger than 2.

C. Relation between pWCET and pET

The relation between the pWCET of a task and the pETs of all jobs of a task is defined using the relation \( \geq \) provided in Definition 3.

Definition 3. [15] Let \( X \) and \( Y \) be two random variables. We say that \( X \) is worse than \( Y \) if \( F_X(x) \leq F_Y(x), \forall x, \) and denote it by \( X \geq Y \).

The pWCET \( C_i \) is an upper bound on the pETs \( C_{i,j}^{\text{pET}}, \forall j \) and it may be described by the relation \( \geq \) as \( C_i \geq C_{i,j}^{\text{pET}}, \forall j \). Graphically \( F_C(.) \) never goes above the curves \( F_{C_{i,j}}(^{\text{pET}})(.), \forall j \). This relation is illustrated in Figure 1. One may notice that there might be two pETs that are not comparable with respect to the relation \( \geq \) defined by Definition 3.

![Fig. 1. Relation between the CDF of the pWCET and the CDFs of the pETs](image)

III. INDEPENDENT TASKS

Since the seminal paper of Liu and Layland [16] the independence of tasks is defined such that the "...requests for a certain task do not depend on the initiation or the completion of requests for other tasks." Moreover the schedulability analysis of independent tasks may be studied under the hypothesis that the tasks do not share any resources except for the processor. In [14] Cucu shows that, thanks to the definition of the pWCET provided by Burns and Edgar [3], this hypothesis is the unique requirement with the respect of the tasks when probabilistic real-time analyses are considered. A probabilistic real-time system is a system with at least one parameter described by a random variable. In this paper tasks have (worst-case) execution times described by random variables.

We present now two problems for task systems with (worst-case) execution times described by random variables:

- the schedulability analysis for fixed-priority scheduling\(^3\) on one processor. This problem allows us to describe the probabilistic independence hypothesis usually required by such analysis. This hypothesis is not stronger than the hypothesis of independent tasks required by deterministic analyses\(^4\).

\(^3\)The fixed-priority scheduling considers that all jobs of a task conserve the same priority during the entire schedule.

\(^4\)We use deterministic analyses as opposite to probabilistic.

- the estimation of the pWCET PF for a task on one processor. This problem allows us to describe the statistical independence hypothesis usually required by such analysis. This hypothesis is not stronger than the hypothesis of independent tasks required by deterministic analyses.

A. Probabilistic independence

We consider \( n \) periodic tasks \( \tau_i \) described by \( (O_i, C_i, D_i, T_i), \forall i = 1, n \) that are scheduled by a preemptive fixed-priority scheduling algorithm on one processor. The task \( \tau_i \) has a higher priority than \( \tau_j, \forall i < j \). The scheduling algorithm is known and given, and proposing one is beyond the purpose of this paper. An interested reader may find a solution to the optimal fixed-priority scheduling algorithms for tasks with worst-case execution times described by random variables in [4].

The response time PF of a job \( \tau_i^j \) is calculated using the operation between two random variables called convolution (defined in Definition 5) that exists if the random variables are (probabilistically) independent as defined in Definition 4.

Definition 4. Two random variables \( X \) and \( Y \) are (probabilistically) independent if they describe two events such that the outcome of one event does not have any impact on the outcome of the other.

Definition 5. The sum \( Z \) of two (probabilistically) independent random variables \( X \) and \( Y \) is the convolution \( X \otimes Y \) where \( P\{Z = z\} = \sum_{k=-\infty}^{\infty} P\{X = k\} P\{Y = z - k\} \).

The calculation of the response time PF of the \( j^{th} \) job of \( \tau_i \) is \( R_i^j \) is provided in [2] as follows

\[
f_{R_i^j} = f_{R_i^j}^{(0, \lambda_{i,j})} + (f_{R_i^j}^{(\lambda_{i,j}, \infty)}) \otimes f_{C_{i,j}}^{\text{pET}}, \quad (4)
\]

where \( \lambda_{i,j} = O_i + jT_i, \forall j \geq 0 \) is the release time of the job \( \tau_i^j \). A solution for Equation (4) may be obtained recursively [2].

Equation (4) may be reformulated as follows:

\[
R_i^j = B_i(\lambda_{i,j}) \oplus \mathcal{I}_i(\lambda_{i,j}) + C_{i,j}^{\text{pET}}, \quad (5)
\]

where \( B_i(\lambda_{i,j}) \) is the accumulated backlog (e.g., the sum of execution times) of higher priority tasks released before \( \lambda_{i,j} \) and still active (not completed yet) at \( \lambda_{i,j} \). \( \mathcal{I}_i(\lambda_{i,j}) \) is the sum of the execution times of higher priority tasks arriving after \( \lambda_{i,j} \).

In Equations (5) and (4) the operation \( \oplus \) indicates that a summation is needed to take into account the execution times of all higher priority tasks active before, at or after \( \lambda_{i,j} \). If the random variables \( C_{i,j}^{\text{pET}}, \forall i \) and \( \forall j \) are (probabilistically) independent, then this operation is the convolution as defined in Definition 5. If the random variables \( C_{i,j}^{\text{pET}}, \forall i \) and \( \forall j \) are (probabilistically) dependent, then the operation \( \oplus \) needs to take into account the dependencies between these random variables. One possible solution is the utilization of copulas [17], but to our best knowledge there is no extension of the results presented by Equations (5) and (4) to the case of copulas. One may note this interesting and yet open problem of probabilistic real-time systems. This open problem may be
related to the study of the impact of (missing) probabilistic independence while convolving as it has been done in [18], but the missing quantification of the number of dependent random variables within this paper does not allow to conclude. Nevertheless one may note this second open problem.

Case of pETs. If the random variables $C_{j}^{i}$ describe pETs of the $j^{th}$ and $\exists j_{1} \neq j_{2}$ such that $C_{j_{1}}^{i} \neq C_{j_{2}}^{i}$, then the random variables are not independent and Equations (5) and (4) cannot be applied with the current knowledge of the literature.

Case of pWCETs. If the random variables $C_{j}^{i}$ describe pWCETs of tasks $\tau_{i}$, then the random variables are independent as they are obtained as upper bounds for pETs. Thus Equations (5) and (4) can be applied by replacing the operation $\oplus$ by $\otimes$ and $C_{j}^{i}$ by $C_{i}$. Equation (4) becomes Equation (6) and Equation (5) becomes Equation (7) as follows (with the same notations as before):

$$f_{R_{i}^{j}} = f_{R_{i}^{j}[0,\lambda_{i,j}]} + (f_{R_{i}^{j}[\lambda_{i,j},\infty]} \otimes f_{C_{i}})$$

$$R_{i}^{j} = B_{i}(\lambda_{i,j}) \otimes I_{i}(\lambda_{i,j}) \otimes C_{i}$$

In conclusion in the case of tasks with pWCETs, the utilization of the definition of pWCET as provided in [3] implies that the (probabilistic) independence of tasks is implicit and it does not require any new hypothesis for a task system with respect to the case of independent tasks described deterministically.

We present in Section III-B what are the hypotheses needed to obtain a pWCET for a task executed on a given platform.

B. Statistical independence

Since the seminal paper of Edgar and Burns [3], different papers [6], [7] have been presented the calculation of pWCETs based on the utilization of extreme value theory (EVT) [19]. The statistical independence required by EVT must be properly checked as underlined by Griffin and Burns [20] and a complete application of EVT is presented in [8].

We present now the extreme value theory and the associated (statistical) independence definition. We end this section by explaining why this (statistical) independence is not a requirement on the top of independence for the tasks of a system.

1) Extreme value theory: Extreme Value Theory (EVT) estimates the probability of occurrence of extreme values which are known to be rare events [3]. More precisely, EVT predicts the distribution function for the maximal values of a set of $n$ observations, which are modelled with random variables. EVT is analogous to Central Limit Theory [21] but instead of estimating the average or the central part of a PF, EVT estimates the extremes [19].

The main result of EVT is provided in Theorem 1 where $F$ denotes the common distribution function of $n$ random variables describing the observations (execution times) of a task $\tau_{i}$ and the pWCET $C_{i} = M_{n}$.

Theorem 1. [19] Let $\{X_{1}, X_{2}, \ldots, X_{n}\}$ be a sequence of independent and identically distributed (i.i.d.) random variables. Let $M_{n} = \max\{X_{1}, X_{2}, \ldots, X_{n}\}$. If $F$ is a non-degenerate distribution function and there exists a sequence of pairs of real numbers $(a_{n}, b_{n})$ such that $a_{n} \geq 0$ and $\lim_{n\to\infty} P(\frac{M_{n} - b_{n}}{a_{n}} \leq x) = F(x)$, then $F$ belongs to either the Gumbel, the Frechet or the Weibull family.

Theorem 1 (as many statistical results) requires that the $n$ random variables $X_{1}, X_{2}, \ldots, X_{n}$ are (probabilistically) independent. Here the variables $X_{1}, X_{2}, \ldots, X_{n}$ describe observations of the execution of task $\tau_{i}$. The independence hypothesis implies that $X_{i}$ is obtained from observed executions that are independent from those contained by $X_{j}$, $\forall i \neq j$. Nevertheless in reality the user of this theory runs the task under study several times and obtains a set of data. In this case in order to fulfill the hypothesis of (probabilistic) independence required by Theorem 1, the user check the (statistical) independence by applying different independence tests on those data [21]. For instance the lag test results may indicate graphically if the data are independent (see Figure 2, right graph) or not (see Figure 2, left graph).

Fig. 2. Graphical results provided by lag test on two sets of observed execution times

In conclusion EVT requires that the observed executions of a task are made independently and this does not imply any extra requirement on the tasks, but on the process of producing and collecting the data. This requirement is also true for another version of extreme value theory called Peaks Over Threshold [22] that may be applied to the problem of estimating pWCETs. To our best knowledge there is no such application of POT to the pWCET estimation and one may note this interesting and yet open problem for probabilistic real-time systems.

These independent executions of tasks may be obtained either by statistical methods [7] or by ensuring specific properties for the platform [23].

IV. RELATED WORK

In this paper we discuss the impact of the independence property on probabilistic systems with (worst-case) execution times of tasks described by random variables. Such systems are the probabilistic systems the most studied among probabilistic real-time systems. These results may be classified in four main types: schedulability analysis, re-sampling, p(WC)ET estimation, and optimal scheduling algorithms.

Schedulability analysis The seminal paper of Lehoczky [1] proposes the first schedulability analysis of task systems with probabilistic execution times. This result and several improvements [24], [25] consider a specific case of PFs for the pETs. Tia et al.[11] and Gardner [12] propose probabilistic analyses for specific schedulers. Abeni et al. [13] proposes probabilistic...
analyses for tasks executed in isolation and a recent work consider time-evolving models [26]. The most general analysis for probabilistic systems with (worst-case) execution times of tasks described by random variables is proposed in [2]. A time-evolving model of pETs is introduced in [27] and an associated schedulability analysis on multiprocessors is presented.

Re-sampling with respect to the p(WC)ETs The schedulability analyses may have an important complexity directly related to the number of possible values of the random variables. This complexity may be decreased by using re-sampling techniques that ensure the safeness (the new response time PF upper bounds the result obtained without re-sampling) [28], [29].

The p(WC)ET estimation Since the seminal paper of Edgar and Burns [3], different papers [6], [7], [20], [8] propose statistical solutions for this problem. To our best knowledge only one paper proposes a pET estimation for a task [9].

Optimal scheduling algorithms For time-evolving pETs of tasks an optimal fixed-priority algorithm is proposed for several processors [30].

To our best knowledge, there is only one paper presenting optimal fixed-priority scheduling algorithms for pWCETs of tasks described by random variables [4]. First surprising result is that algorithms like Rate Monotonic and Deadline Monotonic are no longer optimal. Maxim et al. introduce two definitions of missing the deadline of jobs, respectively, tasks.

Definition 6 (Job deadline miss). For a job \( \tau_{i,j} \) and a priority assignment \( \Phi \), the deadline miss probability \( DMP_{i,j}(\Phi) \) is the probability that the \( j \)-th job of task \( \tau_i \) misses its deadline:

\[
DMP_{i,j}(\Phi) = P(R_{i,j}(\Phi) > D_i).
\]  

Definition 7 (Task deadline miss ratio). For a task \( \tau_i \), a time interval \([a, b]\) and a priority assignment \( \Phi \), the task deadline miss ratio is computed as follows

\[
DMR_{i}(a, b, \Phi) = \frac{P(R_{i}(a, b)(\Phi) > D_i)}{n_{[a,b]}}
\]  

\[
= \frac{1}{n_{[a,b]}} \sum_{j=1}^{n_{[a,b]}} DMP_{i,j}(\Phi),
\]  

where \( n_{[a,b]} = \lceil \frac{b-a}{T} \rceil \) is the number of jobs of task \( \tau_i \) released during the interval \([a, b]\).

With these definitions, the authors propose optimal solutions for:

1) Basic Priority Assignment Problem (BPAP). This problem involves finding a priority assignment such that the DMR of every task does not exceed the threshold specified, i.e. \( DMR_{i}(\Phi) \leq p_i \). Hence, we search for a feasible priority assignment \( \Phi^* \) such that \( DMR_{i}(\Phi^*) \leq p_i \), \( \forall i \).

2) Minimization of the Maximum Priority Assignment Problem (MPAP). This problem involves finding a priority assignment that minimizes the maximum deadline miss ratio of any task. Hence, we search for a priority assignment \( \Phi^* \) such that \( \max_i \{DMR_{i}(\Phi^*)\} = \min_{\Phi} \{\max_i \{DMR_{i}(a, b, \Phi)\}\} \).

3) Average Priority Assignment Problem (APAP). This problem involves finding a priority assignment that minimizes the sum of the deadline miss ratios for all tasks. Hence, we search for a feasible priority assignment \( \Phi^* \) such that \( \sum_i DM R_{i}(a, b, \Phi^*) = \min_{\Phi} \{\sum_i DM R_{i}(a, b, \Phi)\} \).

V. CONCLUSION AND OPEN PROBLEMS

In this paper we study the impact of defining pWCETs of tasks (concept introduced by Burns and Edgar) on the probabilistic real-time systems. This discussion is proposed from the point of view of the notion of independence. Nowadays the probabilistic real-time literature uses three distinct notions of independence: independence of tasks, probabilistic independence among random variables and statistical independence of data. We show that the two later concepts does not imply the loss of the first one. Therefore the probabilistic real-time analyses do not have stronger requirements from the task systems than a deterministic real-time analysis as long as pWCETs are used.

Within this paper we underlined the existence of several open problems interesting for real-time systems with p(WC)ETs:

- the introduction of copulas for systems with dependent pETs;
- the impact of missing probabilistic independence while convolving pETs;
• utilization of extreme value theory for the pWCET estimation while dependences exist.

ACKNOWLEDGMENT

The author would like to thank to all those help with inspiring discussions on probabilistic real-time systems.

REFERENCES