Reprinted from JOURNAL OF COMPUTER AND SYSTEM SCIENCES
All Rights Reserved by Academic Press, New York and London

Vol. 45, No. 3, December 1992
Printed in Belgium

Some Investigations on FCFS Scheduling in Hard Real Time Applications

U. SCHMID AND J. BLIEBERGER

Department of Automation (183/1),
Technical University of Vienna, Treitlstrasse 3, A-1040 Vienna, Austria

Received January 10, 1990; revised April 10, 1991

We investigate some real time behaviour of a (discrete time) single server system with FCFS task scheduling. The main results deal with the probability distribution of a random variable SRD(T), which describes the time the system operates without violating a fixed task service time deadline T. The tree approach used for the derivation of our results is suitable for revisiting problems already solved by queueing theory, too. Relying on a simple general probability model, asymptotic formulas concerning all moments of SRD(T) are determined; for example, the expectation of SRD(T) is proved to grow exponentially in T, i.e., $E[\mathrm{SRD}(T)] \sim C \cdot \kappa^T$ for some $\kappa > 1$. © 1992 Academic Press, Inc.

1. INTRODUCTION

In this paper, we study some aspects concerning the real time behaviour of a discrete time single server system with FCFS task scheduling. Instead of using queueing theory, we apply a special tree approach which is well known from the analysis of data structures, see [KN3; FL1] for a survey. A very complete discussion of queueing theory may be found in [KL1].

The outline of the paper is as follows: After introducing the underlying abstract model and raising some questions of interest, we mention a few real applications. Section 2 contains the description of the probability model forming the basis of our investigations, Section 3 provides the tree approach suitable for the computations in Section 4. Section 5 is devoted to an application, namely a TDMA server with Poisson arrivals. Section 6 concludes the paper with exhibiting some open problems concerning the subject.

We consider a system containing a task scheduler, a task list of finite capacity, and a single server. Tasks arriving at the system are taken by the scheduler and placed into the task list according to the scheduling strategy. The sever always executes the task at the head of the list; thus scheduling is done by rearranging the task list. A dummy task will be generated by the scheduler, if the list becomes empty. If the server executes a dummy task the system is called *idle*, otherwise *busy*.

Rearranging the task list is assumed to occur at discrete points on the time axis

0022-0000/92 $5.00

only, without any overhead. The (constant) time interval between two such points is called a *cycle*. Due to this assumption, we are able to model tasks formed by indivisible (atomic) actions with duration of one cycle. The *task execution time* of a task is the number of cycles necessary for processing the task to completion if it were to occupy the server exclusively. A "regular" task may have an arbitrary task execution time; a dummy task as mentioned above is supposed to consist of a single no-operation action (one cycle). The *service time* of a task is the time (measured in cycles) from the end of the cycle in which the corresponding task arrives at the system to the end of the last cycle of that task.

Obviously, the time axis is covered by *busy periods*, which we assume include the initial idle cycle, too. This definition implies the correspondence between an idle cycle and a busy period with duration of one cycle. A sequence of busy periods without violation of any task's service time deadline followed by a busy period containing at least one deadline violation is called a *run*, the sequence without the last (violating) busy period is referred to by *successful run*.

In order to investigate real time performance, the following random variables are of interest:

(1) The *busy period duration* BPD. This is the time interval (measured in cycles) from the beginning of an idle cycle, in which a task arrival occurs, to the end of the last busy cycle induced, i.e., the length of a whole busy period. We should mention that this duration provides no answer about missing deadlines, since it takes into account the sum of all service times of tasks arriving within the period only, but it should give some insight in system load distributions. BPD is determined by the arrival process only, hence is independent of the scheduling strategy and has been analyzed by classical queueing theory, too, cf. [KL1]. Our analysis is done in another paper, cf. [BS1]; it demonstrates the power of the approach in obtaining the required results quite easily.

(2) The *time to exceed* $\mathrm{TTE}(T)$ and the *successful run duration* $\mathrm{SRD}(T)$. The former is the time interval (measured in cycles) from the beginning of the initial idle cycle to the beginning of the first cycle, causing a fixed task service time deadline of T cycles to exceed, i.e., the time the system operates until the first violation of a task's deadline. $\mathrm{SRD}(T)$ is the time interval from the beginning of the initial idle cycle to the beginning of the (idle) cycle initiating the busy period containing the first violation of a task's deadline T. Obviously, we have $\mathrm{SRD}(T) < \mathrm{TTE}(T)$. In this paper, we restrict ourselves to the investigation of $\mathrm{SRD}(T)$.

Different scheduling strategies may be compared via the distribution of these quantities, even if the arrival process is modeled in a very simple manner (as we did). For example, we may compare the averages of $\mathrm{SRD}(T)$, or the probabilities of finding the system in operation, say, two weeks after power on, without violating a deadline, of course.

Note, that our deadline constraint implies a bounded length of the task list since we suppose FIFO scheduling. In the worst case, a finite capacity task list which is able to hold $T-1$ tasks is sufficient.

According to Section 2, we assume an arrival process, which provides an arbitrary distributed number of task arrivals within a cycle, independent from the arrivals in the preceding cycles and independent from the arbitrary distributed task execution times, too.

To make things clearer, we give a few applications of the above. For example, consider a single processor with a single interrupt line, which executes all machine instructions within a fixed time, a cycle (a few 100 ns, for example). Traditionally, interrupt arrivals will become recognized at the end of an instruction, causing the CPU to process a (reentrant) service routine. An idle cycle corresponds to the execution of an instruction that is not part of an interrupt service routine. Since a cycle is very small, we occasionally may drop the case of more than one interrupt occurrence during a cycle.

A straightforward application is the ordinary FCFS task scheduling problem for a single processor, though it causes some problems is how to justify an equidistant subdivision in atomic actions at a higher level than machine instructions. However, modelling task arrivals by a Poisson process seems to be a possible approach.

Another application of the general model may be found in a server for a TDMA channel (time division multiple access). If we consider a single communication channel shared by multiple (say, n) stations, a common approach for synchronizing transmission activities is TDMA. Each station owns a unique subslot of duration t/n, where it may transmit exclusively (if there are data to transmit, otherwise the subslot is wasted), altogether forming a transmission slot of duration t. Due to the cyclic occurrence of the transmission slot, each station may transmit every t time units. A reasonable order of magnitude for t is 10 ... 100 ms.

To apply our model, we take transmission slots as cycles and assume a constant service time of one cycle, i.e., service corresponds to the transmission of a packet; an idle cycle corresponds to a wasted (sub)slot. The packet arriving process may be modeled by a Poisson process, for example.

2. Probability Model

This section introduces the probability model used for subsequent investigations. We assume arbitrary but independent probability distributions of both the number of task arrivals within a cycle and the task execution times.

The probability generating function (PGF) of the number of task arrivals during a cycle is denoted by

$$A(z) = \sum_{k \geqslant 0} a_k z^k$$

and should meet the constraint $a_0 = A(0) > 0$; i.e., the probability of no arrivals during a slot should be greater than zero. This assures the existence of idle cycles. The definition assumes the independence of arrivals within two arbitrary different cycles.

The PGF of the task execution time (measured in cycles) is denoted by

$$L(z) = \sum_{k \geqslant 0} l_k z^k$$

with the additional assumption $L(0) = 0$; i.e., the task execution time should be greater than or equal to one cycle. Again, this definition assumes task execution times both independent from each other and from the arrival process. Note that we assume an a priori knowledge of the task execution time at the time the task arrives. Since we are studying FCFS scheduling, we may deal with the overall service time, i.e., the number of cycles induced by arrivals within a cycle, instead of using the number of arrivals and corresponding service times separately. Obviously, we obtain

$$P(z) = \sum_{k \geqslant 0} p_k z^k = A(L(z)).$$

In order to justify our computations, we will need some constraints concerning zeros of $P(z) - z$, i.e., fixed points of $P(z)$.

Considering an arbitrary PGF $P(x)$ w.r.t. real arguments x, we obviously state the trivial fixed point $x = 1$. If the Taylor expansion at $x = 1$ exists, valid for x sufficiently large, we have

$$P(x) - x = (x-1)(P'(1) - 1) - R_2(x).$$

Providing the additional assumptions $0 < P'(1) < 1$ and $P''(x) \not\equiv 0$, we obtain for some ε sufficiently small $P(x) - x < 0$ for $x \in (1, 1+\varepsilon)$, since $R_2(x) = O((x-1)^2)$. When x becomes large, $R_2(x)$ increases faster than $(x-1)(P'(1)-1)$ decreases, causing a zero of $P(x) - x$ denoted by κ.

Note, that $P'(\kappa) > 1$, which forces κ to be a simple zero of $P(x) - x$. This is easily proved by applying the first mean value theorem of differential calculus, which states the existence of a v with $\zeta \leqslant v \leqslant \kappa$ and

$$P'(v) = \frac{P(\kappa) - P(\zeta)}{\kappa - \zeta} > \frac{\kappa - \zeta}{\kappa - \zeta} = 1$$

for a $\zeta \in (1, 1+\varepsilon)$. Since $P'(x)$ is monotonic, we obtain $P'(\kappa) \geqslant P'(v) > 1$. The simple zero is easily justified by mentioning the Taylor expansion of $P(x) - x$ at κ.

Considering complex arguments, we show that the trivial zero at $z = 1$ is a simple one and that no other zeros exist within the open disk with radius κ around 0; a radius of convergence $R_P > \kappa$ for $P(z)$ is assumed here. We use the theorem of Rouchè, which states as follows (cf. [CO1]):

THEOREM (Rouchè). *Suppose $f(z)$ and $g(z)$ are meromorphic functions in a neighborhood of the closed disk with radius R around a with no zeros or poles on the* *circle* $\gamma = \{z : |z-a| = R\}$. *If* Z_f, Z_g (P_f, P_g) *are the number of zeros (poles) of* $f(z)$ *and* $g(z)$ *inside* γ, *counted according to their multiplicities, and if*

$$|f(z)+g(z)| < |f(z)| + |g(z)|$$

on γ, *then*

$$Z_f - P_f = Z_g - P_g.$$

We need a weaker, more classical condition only, namely $|f(z)+g(z)| < |f(z)|$. Let $f(z)+g(z) = P(z)$ and $f(z) = z$; hence $g(z) = P(z) - z$. According to our investigations concerning real arguments above, we obtain for any z with $|z| = r < \kappa$,

$$|f(z)+g(z)| = |P(z)| \leqslant P(|z|) = P(r) < r = |z| = |f(z)|,$$

which establishes the conditions of the theorem. Note, that this inequality ensures that no zeros of $g(z) = P(z) - z$ on $|z| = r$ exist; the analyticity of both $f(z)$ and $g(z)$ excludes poles on $|z| = r$. Thus, $g(z)$ has exactly as many zeros as $f(z)$, i.e., exactly one simple zero within the disk of radius $r < \kappa$ around 0. On $|z| = \kappa$, we obtain a second zero of $P(z) - z$ at $z = \kappa$ and no others. Relying on these results, we are able to state the required conditions as follows:

Let $P(z)$ denote the PGF of the number of cycles induced by arrivals within a cycle, which should meet the following constraints:

(1) The average number of cycles induced by arrivals within a cycle should be smaller than one, i.e., $0 < P'(1) < 1$. Since we investigate real time applications, the case of average high load $(P'(1) \approx 1)$ seems to be of no concern. Note, that this assumption implies $p_0 = P(0) > 0$, since $1 > P'(1) \geqslant P(1) - p_0 = 1 - p_0$.

(2) $P''(z) \not\equiv 0$, i.e., we explicitly exclude the trivial case $P(z) = p_0 + (1-p_0)z$.

(3) The radius of convergence R_P of $P(z)$ should be sufficiently large, such that some $\kappa < R < R_P$ may be determined with the property that $P(z) - z$ has only its real, simple zeros $z = 1$ and $z = \kappa$ within the closed disk with radius R around 0. In order to justify some remainder terms in Section 4, we should in fact choose $R < \max(R_P, \kappa^2)$. Note that this condition forces all moments of $P(z)$ to be finite since $R_P > \kappa > 1$.

We should mention that the number of probability distributions meeting our constraints is considerably limited due to the required independency. An example for a suitable model is based on an interarrival distribution with the so-called memoryless property, i.e., an exponential or geometric distribution, leading to Poisson- or Bernoulli-type arrivals within a cycle, respectively.

3. Tree Approach

We start our treatment by introducing an *arrival sequence* $\{a_n\}$, $n \geqslant 0$, where $a_n \geqslant 0$ counts the number of cycles caused by task arrivals during the nth busy cycle following the initial (idle!) cycle. We will establish a one-to-one mapping between arrival sequences and the family of planted planar trees, which provides a nice correspondence between deadline constraints and limited widths of the tree. Due to this fact, we may relate the original problem of investigating the random variable $\mathrm{SRD}(T)$ to a counting problem regarding a special (sub)family $\mathscr{B}_T$ of trees. Let us start with an example; consider the arrival sequence

$$(3, 2, 0, 0, 0, 1, 2, 0, 0)$$

and the corresponding tree shown in Fig. 1. Each vertex corresponds to a cycle n; the number of successors of a vertex equals a_n, the number of (busy) cycles caused by arrivals during the cycle; the root corresponds to the initial idle cycle 0. The execution sequence is related to the preorder traversal policy (left to right) of the tree. The "aligned" representation of the tree above will help us in establishing the deadline property mentioned before.

For convenience, each vertex is labeled by an expanded string representation of the actual task list at the beginning of the corresponding cycle, i.e., by all cycles currently forming the task list. The kth cycle of the nth task is denoted by n_k. New cycles are attached at the end of the string, the cycle actually executed is removed at the front of it. Note, however, that construction and reconstruction of tree and arrival sequence, respectively, does not depend on this labeling.

Looking carefully at our example, one obtains that the number of cycles forming the task list for all vertical aligned vertices is equal; and this is in fact true for all such trees due to the construction principle. This number represents the time interval (measured in cycles) until completion of the last cycle in the list; hence limiting

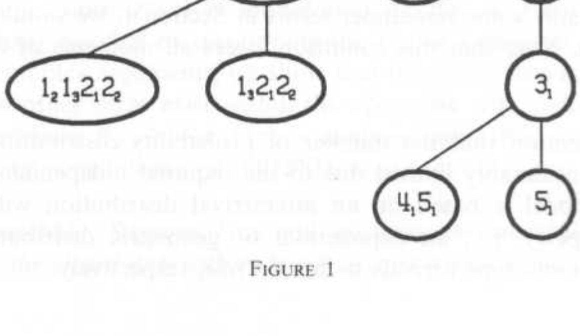


Figure 1

the service times of the tasks by a deadline T is reflected by limiting the width of the tree to T vertices!

To obtain the connection with our probability model, we simply have to attach weights to all vertices, equal to the probability of having their specific number of successors. The ordinary generating function (OGF) of this special family $\mathscr{B}_T$ of trees is the PGF of the length of a busy period conditioned by the fact that the busy period contains no deadline violation.

4. Moments of SRD(T)

As mentioned in Section 1, a run denotes a sequence of busy periods not violating any task's deadline followed by a busy period with at least one deadline violation. Let

$$b_{k,T} = \text{prob}\{\text{A nonviolating busy period of length } k \text{ cycles occurs}\}$$

and

$$B_T(z) = \sum_{k \geq 0} b_{k,T} z^k$$

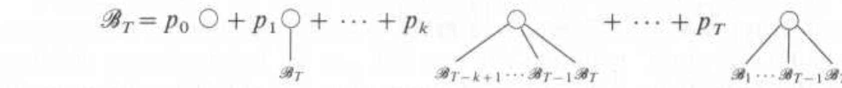
be the corresponding PGF. The PGF of the random variable SRD(T), i.e., the length of a successful run, is given by

$$S_T(z) = \sum_{k \geq 0} s_{k,T} z^k = \frac{1 - B_T(1)}{1 - B_T(z)}. \tag{0}$$

This follows from the fact that the PGF of the length of an arbitrary number of nonviolating busy periods is $\sum_{n \geq 0} B_T(z)^n$ and that the probability of the occurrence of the terminating violation busy period equals $1 - B_T(1)$.

In order to derive $B_T(z)$, we start with the following symbolic equation concerning our family of width-constrained trees $\mathscr{B}_T$. This family appears in the analysis of a simple register function regarding T-ary operations, too; cf. [KP1; FL2] for details. In fact, there is a relation to the so-called left-sided height of a tree.

With p_k denoting the probability of obtaining k cycles induced by arrivals within a cycle (cf. Section 2), we have



for all $T \geq 1$. According to [FL3], this symbolic equation translates into a recurrence relation of the ordinary generating function

$$B_T(z) = \sum_{k=0}^{T} p_k z \prod_{j=T-k+1}^{T} B_j(z),$$

since each vertex with k successors is weighted by $p_k z$; the coefficient of z^n in $B_T(z)$, denoted by $b_n = [z^n]\, B_T(z)$, is the probability of obtaining a tree with exactly n vertices. Defining

$$Q_n(z) = \frac{1}{B_n(z) \cdots B_1(z)}$$

$$Q_0(z) = 1$$

and the corresponding bivariate generating function

$$Q(s, z) = \sum_{k \geq 0} Q_k(z)\, s^k,$$

we obtain

$$B_T(z) = \frac{Q_{T-1}(z)}{Q_T(z)}.$$

Multiplying our fundamental recurrence relation by $Q_T(z)$ yields

$$Q_{T-1}(z) = z \sum_{k=0}^{T} p_k Q_{T-k}(z).$$

Multiplying both sides by s^T and summing up for $T \geq 1$, we find

$$Q(s, z) = \frac{zp_0}{zP(s) - s}.$$

We should mention that, by a simple expansion of the bivariate generating function, $Q_T(z)$ is proved to be a polynomial of degree T in $1/z$; all coefficients are explicitly expressible in terms of p_k. This is easily seen by rewriting

$$Q(s, z) = \frac{zp_0}{s} \sum_{n \geq 1} \left(\frac{s}{zP(s)} \right)^n;$$

hence

$$Q_T(z) = p_0 \sum_{n=0}^{T} [s^n] \left(\frac{1}{P(s)} \right)^{T-n+1} z^{-(T-n)}.$$

Note, that the restriction of the range of summation is justified by the property $P(0) > 0$, according to our contraints mentioned in Section 2, since $s/P(s) = w_1 s + w_2 s^2 + \cdots$ and $w_1 = 1/P(0)$.

Fortunately, the bivariate generating function $Q(s, z)$ enables us to use singularity analysis techniques for obtaining results concerning $Q_T(z)$ and $B_T(z)$; hence we are not forced to make use of explicit expressions. Note, however, that

$B_T(z)$ is a rational function. Quantities related to $Q_T(1)$ arise frequently in the investigation of the maximum of a sum of independent random variables, cf. [TA1] for details.

We will determine the mth derivative of $Q_T(z)$, denoted by $Q_T^{(m)}(z)$, evaluated at the point $z=1$. For practical applications, the deadline T of a task should be large compared to the duration of a cycle; hence asymptotic results for large T are satisfactory. We easily obtain

$$Q_T^{(m)}(1)=Q_T^{(m)}(z)|_{z=1}=m!\,[(z-1)^m][s^T]\,Q(s,z).$$

The expansion of $Q(s,z)$ at $z=1$ is found by mentioning that

$$Q(s,z)=\frac{zp_0}{zP(s)-s}$$

$$=-\frac{p_0}{P(s)}\cdot\frac{(z-1)\dfrac{P(s)}{s-P(s)}}{1-(z-1)\dfrac{P(s)}{s-P(s)}}-\frac{p_0}{s-P(s)}\cdot\frac{1}{1-(z-1)\dfrac{P(s)}{s-P(s)}};$$

hence we are able to pick up the coefficient of $[(z-1)^m]$ directly by using the geometric series. For $m\geqslant 1$, we obtain

$$[(z-1)^m]\,Q(s,z)=-\frac{p_0 s(P(s))^{m-1}}{(s-P(s))^{m+1}}.$$

For $m=0$, we have

$$[(z-1)^0]\,Q(s,z)=Q(s,1)=-\frac{p_0}{s-P(s)}.$$

According to methods from singularity analysis, the coefficient of s^T is mainly determined by the singularity at $s=1$, resulting from the denominator vanishing at this point, cf. Section 2. An overview to asymptotic methods, especially concerning the method of Darboux, may be found in [FL1; BE1]. However, we will need elementary techniques only, namely a weaker version of the so-called Cauchy's estimates.

Conventionally, we write $f(x)=O(g(x))$ for $x\to x_0$, if there exists some real constant $M>0$ independent of x which guarantees $|f(x)|\leqslant M\,|g(x)|$ for all x in a suitable neighborhood of x_0. We use the notation $f(x)\sim g(x)$ for $x\to x_0$, if $\lim_{x\to x_0}f(x)/g(x)=1$.

THEOREM (Cauchy's Estimates). *Suppose that $A(z)=\sum_{n\geqslant 0}a_n z^n$ has a radius of convergence $r>0$, and let $R<r$ denote an arbitrary real, positive number. We have*

$$a_n=O(R^{-n}).$$

Since we need more exact asymptotic expansions for $m=0$ and $m=1$, we treat them first. Expanding $P(s)$ in powers of $s-1$ yields

$$P(s)=1+P'(1)(s-1)+\frac{P''(1)}{2}(s-1)^2+O((s-1)^3) \qquad \text{for} \quad s\to 1,$$

the remainder represents a function, say $T(s)$, with a triple zero at $s=1$. Thus,

$$s-P(s)=(1-P'(1))(s-1)\left(1-\frac{P''(1)}{2(1-P'(1))}(s-1)+O((s-1)^2)\right). \tag{1}$$

Following this, we are able to conclude the existence of a function $R(s)=1+O(s-1)$ which fulfills

$$\frac{1}{s-P(s)}=\frac{R(s)}{(1-P'(1))(s-1)};$$

hence we find

$$\begin{aligned} Q(s,1)&=\frac{-p_0}{1-P'(1)}\cdot\frac{R(s)}{s-1}\\ &=\frac{-p_0}{1-P'(1)}\cdot\frac{1}{s-1}+W(s). \end{aligned}$$

Note that $W(s)=O(1)$ for $s\to 1$, i.e., has no singularity at $s=1$. According to

$$\begin{aligned} [s^T]\,Q(s,1)&=[s^T]\,\frac{-p_0}{1-P'(1)}\cdot\frac{1}{s-1}+[s^T]\,W(s)\\ &=\frac{p_0}{1-P'(1)}+[s^T]\,W(s), \end{aligned}$$

we need the coefficient $[s^T]\,W(s)$, which is determined by the singularities of $W(s)$; since the dominant term $1/(s-1)$ is analytic for all $s\neq 1$, we have to take into account the singularities of $Q(s,1)$ for $s>1$.

According to Section 2, we have an additional simple polar singularity at $\kappa>1$ within the closed disk with radius R around 0. Expanding $P(s)$ yields

$$P(s)=\kappa+P'(\kappa)(s-\kappa)+O((s-\kappa)^2) \qquad \text{for} \quad s\to\kappa;$$

hence

$$s-P(s)=(1-P'(\kappa))(s-\kappa)+O((s-\kappa)^2).$$

Thus, in a neighborhood of κ, $Q(s,1)$ and hence $W(s)$ fulfills

$$W(s)=\frac{p_0}{\kappa(1-P'(\kappa))}\cdot\frac{1}{1-s/\kappa}+O(1).$$

The coefficient of s^T evaluates to

$$[s^T]\, W(s) = \frac{p_0}{\kappa(1 - P'(\kappa))}\, \kappa^{-T} + O(R^{-T});$$

the remainder follows from Cauchy's estimates by mentioning the fact that the function represented by $O(1)$ has no singularities within the closed disk of radius R around 0. Remembering that $P'(\kappa) > 1$ according to Section 2, we finally obtain

$$Q_T(1) = \frac{p_0}{1 - P'(1)} - \frac{p_0}{\kappa(P'(\kappa) - 1)}\, \kappa^{-T} + O(R^{-T}).$$

In order to investigate the case $m = 1$, we find by using Eq. (1) that

$$\frac{1}{(s - P(s))^2} = \frac{1}{(1 - P'(1))^2} \cdot \frac{1}{(s-1)^2} + \frac{P''(1)}{(1 - P'(1))^3} \cdot \frac{1}{s-1} + O(1).$$

This is justified by using the geometric series $1/(1-x) = 1 + x + O(x^2)$. We obtain

$$\begin{aligned} Q_T^{(1)}(1) &= -p_0 [s^{T-1}]\, \frac{1}{(s - P(s))^2} \\ &= -\frac{p_0 T}{(1 - P'(1))^2} + \frac{p_0 P''(1)}{(1 - P'(1))^3} + O(\kappa^{-T}). \end{aligned}$$

The remainder is justified in analogy to the considerations regarding the case $m = 0$ above. $1 < \kappa < R$ denotes the (real) polar singularity of $1/(s - P(s))$. Obviously the coefficients resulting from the fractional terms $1/(s-1)^k$ are their Taylor coefficients when expanding at $s = 0$.

For the general case ($m \geqslant 0$), the previous investigations enable us to conclude the existence of functions $R_m(s) = 1 + O(s-1)$ which fulfill

$$\frac{1}{(s - P(s))^{m+1}} = \frac{R_m(s)}{(1 - P'(1))^{m+1} (s-1)^{m+1}}.$$

Thus, we obtain

$$Q_T^{(m)}(1) = -m!\, p_0 [s^{T-1}]\, \frac{P(s)^{m-1}}{(s - P(s))^{m+1}} = O(T^m).$$

Mentioning $P(s) = 1 + O(s-1)$, we have no contributions from $P(s)^{m-1}$; and, for the same reason, from $R_m(s)$, too. Note, that the remainder is not uniform in m, i.e., only valid for m fixed. We summarize the considerations above in the following lemma.

LEMMA 1. *With the notations above, the* m*th derivative* (m *arbitrary but fixed*) *of* $Q_T(z)$ *evaluated at* $z = 1$ *fulfills*

$$Q_T^{(m)}(1) = O(T^m).$$

More accurate asymptotic expansions for $m = 0$ *and* $m = 1$ *are*

$$Q'_T(1) = -\frac{p_0 T}{(1 - P'(1))^2} + \frac{p_0 P''(1)}{(1 - P'(1))^3} + O(\kappa^{-T})$$

$$Q_T(1) = \frac{p_0}{1 - P'(1)} - \frac{p_0}{\kappa(P'(\kappa) - 1)} \kappa^{-T} + O(R^{-T}).$$

Now we are able to return to the PGF of SRD(T), which has been evaluated to

$$S_T(z) = \sum_{k \geq 0} s_{k,T} z^k = \frac{1 - B_T(1)}{1 - B_T(z)},$$

cf. Eq. (0). We investigate the moments of this distribution, i.e., the quantities

$$E^n(T) = \mathrm{E}[\mathrm{SRD}(T)^n] = \sum_{k \geq 0} k^n s_{k,T}.$$

In addition, we define the nth factorial moment by

$$F^n(T) = \sum_{k \geq 0} [k]_n s_{k,T} = S_T^{(n)}(1),$$

where $[k]_n = k(k-1) \cdots (k-n+1)$ denotes the falling factorial. Note that n is assumed to be fixed; all $O(\ )$-terms are uniform in T only. Since $[k]_n = k^n + O(k^{n-1})$, we obtain

$$F^n(T) = E^n(T) + O(E^{n-1}(T)).$$

If we could provide $F^{n-1}(T) = O(F^n(T))$, a simple induction argument would show that

$$E^n(T) = F^n(T) + O(F^{n-1}(T)); \tag{2}$$

hence it seems reasonable to investigate the factorial moments. Since

$$B_T(z) = Q_{T-1}(z)\, Q_T^{-1}(z),$$

we have

$$S_T(z) = g(B_T(z))$$

for $g(z)=(1-B_T(1))/(1-z)$. An easy computation shows

$$g^{(j)}(z)|_{z=B_T(1)}=g^{(j)}(B_T(1))=\frac{j!}{(1-B_T(1))^j}$$

for all $j\geqslant 0$. Using the formula of Faà di Bruno (cf. [KN1, p. 50])

$$(b(a(z)))^{(n)}|_{z=t}$$
$$=\sum_{j=0}^{n} b^{(j)}(a(t)) \sum_{\substack{k_1+k_2+\cdots+k_n=j\\ k_1+2k_2+\cdots+nk_n=n\\ k_i\geqslant 0}} \frac{n!}{k_1!\,(1!)^{k_1}\cdots k_n!\,(n!)^{k_n}}(a^{(1)}(t))^{k_1}\cdots(a^{(n)}(t))^{k_n},$$

we are able to express $S_T^{(n)}(1)$ in terms of $g^{(j)}(B_T(1))$ and $B_T^{(j)}(1)$; setting $b(z)=g(z)$, $a(z)=B_T(z)$, and $t=1$, we find

$$S_T^{(n)}(1)=\sum_{j=0}^{n}\frac{1}{(1-B_T(1))^j}\sum_{\substack{k_1+\cdots+k_n=j\\ k_1+\cdots+nk_n=n\\ k_i\geqslant 0}} c_{j,n,k_1,k_2,\ldots,k_n}(B_T^{(1)}(1))^{k_1}\cdots(B_T^{(n)}(1))^{k_n}$$

with the abbreviation

$$c_{j,n,k_1,k_2,\ldots,k_n}=j!\,\frac{n!}{k_1!\,(1!)^{k_1}\cdots k_n!\,(n!)^{k_n}}.$$

According to the formula of Leibniz, we obtain

$$B_T^{(i)}(z)=\sum_{m=0}^{i}\binom{i}{m}\left(\frac{1}{Q_T(z)}\right)^{(m)}Q_{T-1}^{(i-m)}(z).$$

In order to find the mth derivative of $Q_T^{-1}(z)$, we make use of Faà di Bruno's formula again. Temporarily setting $a(z)=Q_T(z)$ and $b(z)=1/z$ yields

$$\left.\left(\frac{1}{Q_T(z)}\right)^{(m)}\right|_{z=1}$$
$$=\sum_{j=0}^{m}\frac{1}{Q_T^{j+1}(1)}\sum_{\substack{k_1+\cdots+k_m=j\\ k_1+\cdots+mk_m=m\\ k_i\geqslant 0}} d_{j,m,k_1,k_2,\ldots,k_m}(Q_T^{(1)}(1))^{k_1}\cdots(Q_T^{(m)}(1))^{k_m}.$$

But, mentioning Lemma 1, we have $Q_T^{(i)}(1)=O(T^i)$; hence the inner sum provides an overall contribution of $O(T^{k_1+2k_2+\cdots+mk_m})=O(T^m)$. Moreover, according to this lemma, we have $Q_T(1)=p_0/(1-P'(1))+O(\kappa^{-T})$, which yields $Q_T^{-1}(1)=O(1)$; therefore,

$$\left.\left(\frac{1}{Q_T(z)}\right)^{(m)}\right|_{z=1}=O(T^m).$$

Substituting these asymptotic expansions in the formula above, we obtain by similar reasoning as before

$$B_T^{(i)}(1) = O(T^i).$$

Using this in our first application of Faà di Bruno's formula, an overall contribution of the inner sum equal to $O(T^n)$ is found. Because of

$$1 - B_T(1) = 1 - \frac{Q_{T-1}(1)}{Q_T(1)}$$
$$= \frac{\kappa^{-T}}{c}(1 + O((R/\kappa)^{-T}))$$

with abbreviations resulting from Lemma 1

$$a = \frac{p_0}{1 - P'(1)}$$
$$b = \frac{p_0}{\kappa(P'(\kappa) - 1)}$$
$$c = \frac{a}{b(\kappa - 1)} = \frac{\kappa(P'(\kappa) - 1)}{(\kappa - 1)(1 - P'(1))},$$

the major contributions come from $(1 - B_T(1))^{-j}$ with $j = n$ and

$$\frac{1}{(1 - B_T(1))^n} = c^n \kappa^{nT}(1 + O((R/\kappa)^{-T})).$$

Note, that we should choose $R < \kappa^2$ in order to justify our remainder. However, we may discard all terms of the outer sum concerning $S_T^{(n)}(1)$, except for $j = n$; i.e., we obtain

$$S_T^{(n)}(1) = \sum_{\substack{k_1 + \cdots + k_n = n \\ k_1 + \cdots + nk_n = n \\ k_i \geqslant 0}} \frac{c_{n,n,k_1,k_2,\ldots,k_n}}{(1 - B_T(1))^n} (B'_T(1))^{k_1} \cdots (B_T^{(n)}(1))^{k_n} + O(T^n \kappa^{(n-1)T})$$
$$= c_{n,n,n,0,\ldots,0} \frac{(B'_T(1))^n}{(1 - B_T(1))^n} + O(T^n \kappa^{(n-1)T})$$
$$= n! \left(\frac{B'_T(1)}{1 - B_T(1)} \right)^n + O(T^n \kappa^{(n-1)T}),$$

since the conditions concerning the inner sum hold for $k_1 = n$ only. Substituting the expansion above, we find

$$S_T^{(n)}(1) = n!\, c^n \kappa^{nT} (B'_T(1))^n (1 + O((R/\kappa)^{-T})),$$

the old remainder disappears within the new one. The last task is the evaluation of

$$B'_T(1) = \frac{Q'_{T-1}(1)}{Q_T(1)} - \frac{Q_{T-1}(1)\, Q'_T(1)}{Q_T^2(1)}.$$

According to Lemma 1, we have $Q_T(1) = p_0/(1 - P'(1)) + O(\kappa^{-T})$; thus it is easy to find $Q_T^{-1}(1) = (1 - P'(1))/p_0 + O(\kappa^{-T})$. With $Q'_T(1)$ from the same lemma, we obtain

$$B'_T(1) = \frac{1}{1 - P'(1)} + O(T\kappa^{-T}).$$

We summarize these results concerning the Taylor expansion of $B_T(z)$ at $z = 1$ in the following lemma.

LEMMA 2. *With the notations above, the first few coefficients in the Taylor expansion of $B_T(z)$ at $z = 1$ are*

$$B_T(1) = 1 - \frac{(\kappa - 1)(1 - P'(1))}{\kappa(P'(\kappa) - 1)} \kappa^{-T} + O(R^{-T})$$

$$B'_T(1) = \frac{1}{1 - P'(1)} + O(T\kappa^{-T}).$$

This completes our computations concerning the factorial moments of SRD(T). We have

$$F^n(T) = n!\, \mu(T)^n (1 + O((R/\kappa)^{-T}))$$

with

$$\mu(T) = \kappa^T \frac{\kappa(P'(\kappa) - 1)}{(\kappa - 1)(1 - P'(1))^2}.$$

Since $F^{n-1}(T) = O(F^n(T))$, the condition for Eq. (2) is justified and we obtain

$$E^n(T) = F^n(T) + O(F^{n-1}(T)).$$

The remainder above disappears within the remainder term established for $F^n(T)$, as can be shown by straightforward estimations using $R < \kappa^2$; hence our final result follows:

THEOREM 1. *With the notations above, the* n*th moment* (n *arbitrary but fixed*) *of* SRD(T) *fulfills*

$$E^n(T) = n!\, \mu(T)^n (1 + O((R/\kappa)^{-T}))$$

with

$$\mu(T) = \kappa^T \frac{\kappa(P'(\kappa) - 1)}{(\kappa - 1)(1 - P'(1))^2}.$$

5. POISSON ARRIVALS

This section deals with the application of the preceding general formulas to the Poisson case, which fits our TDMA example of the first section. Relying on these results, we may compare different scheduling techniques w.r.t. their behaviour concerning missing a deadline. Suppose

$$P(z) = e^{\lambda(z-1)},$$

the PGF of a Poisson distribution with rate $0 < \lambda < 1$. Note, that $P'(1) = \lambda$, i.e., the rate equals the average number of cycles induced by arrivals within a cycle, and that we are mainly interested in small values of λ, cf. Section 2.

In order to determine the most critical quantity κ, we have to compute the (unique) real solution $\kappa > 1$ of

$$P(s) - s = 0.$$

Straightforward manipulations show that solutions of the above are obtained by investigating the solutions of

$$ze^{-z} = \mu,$$

where $z = \lambda s$ and $\mu = \lambda e^{-\lambda} < 1$ instead. This is done according to [DB1, pp. 25–28]; by using the substitution

$$z = -\log \mu + \log(-\log \mu) + w,$$

we obtain an equation in w,

$$e^w - 1 + \frac{w}{\log \mu} + \frac{\log(-\log \mu)}{\log \mu} = 0.$$

Introducing the abbreviations

$$v = (-w)$$
$$\sigma = 1/\log \mu$$
$$\tau = \log(-\log \mu)/\log \mu,$$

we have

$$e^{-v}-1-\sigma v+\tau=0.$$

As shown in [DB1, p. 27], if $|\sigma|$, $|\tau|$ sufficiently small, there exists a unique solution in a domain $|v|<b$ for some $b>0$, which can be written as an absolutely convergent power series

$$v=\sum_{k\geqslant 0}\sum_{m\geqslant 0} c_{km}\sigma^k\tau^{m+1}$$

with some constants c_{km}. Since $\mu=O(\lambda)$ for $\lambda\to 0$, both σ and τ become arbitrarily small for λ sufficiently small. Obviously, we have the asymptotic expansion

$$v=O\left(\frac{\log(-\log\mu)}{-\log\mu}\right);$$

hence we obtain

$$z=-\log\mu+\log(-\log\mu)+O\left(\frac{\log(-\log\mu)}{-\log\mu}\right).$$

Mentioning

$$-\log\mu=-\log\lambda+\lambda$$

and

$$\log(-\log\mu)=\log(-\log\lambda)+O(\lambda/-\log\lambda),$$

where we used $\log(1\pm x)=O(x)$ for small x, we find

$$z=-\log\lambda+\log(-\log\lambda)+O(\log(-\log\lambda)/-\log\lambda).$$

Note, that the remainder causes the term λ to vanish. Remembering the fact that $\kappa=s=z/\lambda$, we obtain

$$\kappa=\frac{-\log\lambda}{\lambda}+\frac{\log(-\log\lambda)}{\lambda}+O\left(\frac{\log(-\log\lambda)}{\lambda(-\log\lambda)}\right),$$

and this is in fact the required solution since $\kappa>1$ for λ sufficiently small. However, numerical computations show that the approximation of κ (the first two terms in the asymptotic expansion) is satisfactory for very small $\lambda<0.1$ only.

λ	κ
0.1	37.15
0.2	14.30
0.3	7.88
0.4	5.05
0.5	3.51
0.6	2.58
0.7	1.97
0.8	1.54
0.9	1.23

TABLE I

Finally, we will give some numerical results concerning our TDMA model. We assume a transmission slot, i.e., a cycle duration of 10 ms, deadlines ranging from 10 to 100 cycles (0.1 s to 1 s), and input arrival rates from 0.1 to 0.9 arivals/cycle (10 to 90 arrivals/s). Table I shows the values of κ w.r.t. different input rates. Figure 2 shows the expectation of $\mathrm{SRD}(T)$ versus the deadline T, which is equal to the standard deviation, too. The y-axis is log-scaled (10^2 cycles per division), the x-axis is linear. Note, that a second corresponds to 10^2 cycles; one year is approximately 3×10^9 cycles.

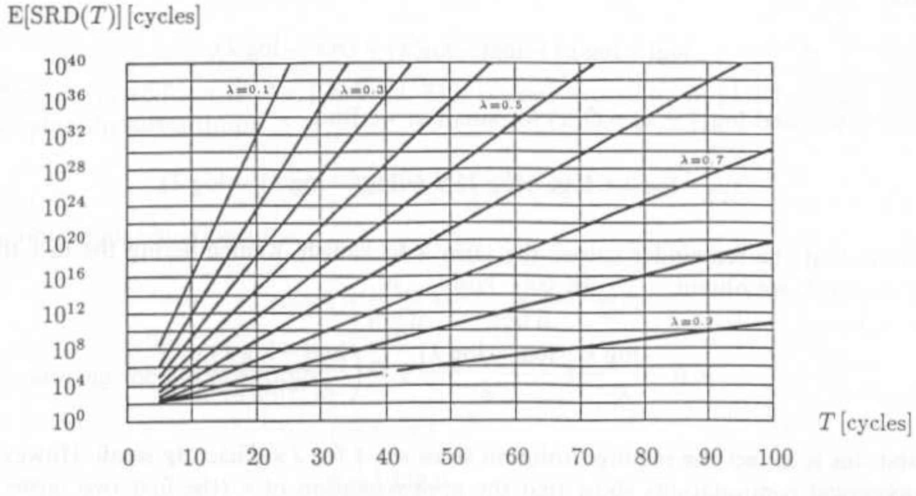


FIGURE 2

6. Conclusions

Due to our intention to study the system with regard to its real time behaviour, the various results obtained by classical queueing theory are not useful for our purposes. For example, performance engineering results traditionally make use of convenient equilibrium assumptions, which are justified by stable operating conditions. Instead, we have examined the system with respect to its ability to meet the deadlines of *all* tasks arriving at the system from the time it is turned on to the year 9999. We have found impressive results concerning the expectation of SRD(T) (unfortunately, weakened by a large standard deviation) and even the general results show very nice deadline meeting properties. Though they are mainly caused by our somewhat stationary probability model, they are still useful because of their non-equilibrium nature. For example, if arrival probability distributions concerning stress situations are available, we could determine some limits regarding the tolerable duration of such stress periods.

The comparison of our results concerning the Poisson case (cf. Section 6) with the results from the analysis of preemptive LCFS scheduling (last come first served, cf. [BS1]) shows that FCFS provides a significantly better behaviour w.r.t. missing deadlines, especially for low and medium rates.

The very detailed computations contained in the preceding sections are primarily addressed to the mathematically inclined reader, who (hopefully) will find them relatively straightforward. Note, however, that a simulation approach concerning SRD(T) for reasonable values of T seems to be impossible, even on a CRAY computer, cf. our numerical results. Thus, we have solved a problem by means of analytic modelling, which is not tractable by simulation, providing a counter-example to the widespread view of simulation being a panacea.

Needless to say, this approach is only a modest start to analytic modelling of systems for real time applications; there are a lot of more or less important problems left to the reader: It seems necessary to define and investigate other quantities describing real time behaviour better than our SRD(T) does, for all possible scheduling techniques, of course. Further, releasing the fixed deadline assumption, adding system overhead for scheduling and dispatching, dropping the limitation to a single server and covering the occurrence of deterministic and cyclically created tasks are of special interest. Minor modifications of our model to meet special applications are often straingtforward.

Obviously, a crucial point is how to model the task arrival process to meet practical requirements. This problem, which is central to all attempts of analytic modelling a real application, is not solved sufficiently. In order to preserve the tractability of the computations, one is traditionally forced to use the well-thumbed exponential or geometric distributions, or parameter variant normal distributions as in diffusion approximation. Unfortunately, these approaches are justified for some traditional applications only (large timesharing systems, for example), but it seems to be unlikely they are successful in real time applications.

Hence, the development of an approach which allows the extension of our stationary probability model to a more suitable dynamic one seems to be of central importance for analytic modelling of hard real time systems. In order to obtain an adequate model, it is important to investigate applications with regard to the stimuli they are concerned with, i.e., there is a need of know-how in monitoring a technical process; both how to do it and what quantities are to be monitored to obtain the desired characteristics.

On the other hand, refined techniques for tracting the theoretical part are necessary in order to make use of an adequate model. Really, a lot of theoretical and practical work remains to be done!

References

[BE1] E. A. Bender, Asymptotic methods in enumeration, *SIAM Rev.* **16** (1974), 485–515.

[BS1] J. Blieberger and U. Schmid, Some investigations on preemptive LCFS scheduling in hard real time applications, to appear in Performance Evaluation.

[CO1] J. B. Conway, "Functions of one Complex Variable," 2nd ed., Springer-Verlag, New York, 1978.

[DB1] N. G. de Bruijn, "Asymptotic Methods in Analysis," Dover, New York, 1981.

[KP1] P. Kirschenhofer and H. Prodinger, On the recursion depth of special tree traversal algorithms, *Inform. and Comput.* **74**, No. 1 (1987), 15–32.

[KL1] L. Kleinrock, "Queueing Systems," Vols. 1, 2, Wiley, New York, 1975.

[FL1] J. S. Vitter and Ph. Flajolet, "Handbook of Theretical Computer Science," North-Holland, Amsterdam, 1989.

[FL2] Ph. Flajolet, J. C. Raoult, and J. Vuillemin, The number of registers required to evaluate arithmetic expressions, *Theoret. Comput. Sci.* **9** (1979), 99–125.

[FL3] Ph. Flajolet, "Analyse d'algorithmes de manipulation d'arbres et de fichiers," pp. 34–35, Cahiers de BURO, Brussels, 1981.

[KN1] D. E. Knuth, "The Art of Computer Programming. Vol. 1. Fundamental Algorithms," 2nd ed., Addison–Wesley, Reading, MA, 1973.

[KN3] D. E. Knuth, "The Art of Computer Programming. Vol. 3. Sorting and Searching," 2nd ed., Addison–Weley, Reading, MA, 1973.

[TA1] L. Takács, "Combinatorial Methods in the Theory of Stochastic Processes," Krieger, Huntington, New York, 1977.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium