THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Parallel computing, failure recovery, and extreme values

Lars Nørvang Andersen and Søren Asmussen

Research Report

No. 13 | August 2007

Parallel computing, failure recovery, and extreme values

This Thiele Research Report is also Research Report number 500 in the Stochastics Series at Department of Mathematical Sciences, University of Aarhus, Denmark.

Parallel computing, failure recovery, and extreme values *

Lars Nørvang Andersen
Søren Asmussen

Department of Mathematical Sciences, Aarhus University,
Ny Munkegade, DK-8000 Aarhus C, Denmark
{larsa,asmus}@imf.au.dk

August 7, 2007

Abstract

A task of random size T is split into M subtasks of lengths $T_1, \ldots, T_M$, each of which is sent to one out of M parallel processors. Each processor may fail at a random time before completing its allocated task, and then has to restart it from the beginning. If $X_1, \ldots, X_M$ are the total task times at the M processors, the overall total task time is then $Z_M = \max_{1,\ldots,M} X_i$. Limit theorems as $M \to \infty$ are given for Z_M, allowing the distribution of T to depend on M. In some cases the limits are classical extreme value distributions, in others they are of a different type.

Key words Cramér-Lundberg approximation, failure recovery, Fréchet distribution, geometric sums, Gumbel distribution, heavy tails, logarithmic asymptotics, mixture distribution, power tail, RESTART, triangular array

1 Introduction

Consider a job that ordinarily would take a time T to be executed on some system (e.g., CPU). If at some time $U < T$ the processor fails, the job may take a *total time* $X \geq T$ to complete. We let F, G be the distributions of T, U with $H = H_{F,G}$ the distribution of X, which in addition to F, G will depend on the failure recovery scheme.

Many papers discuss methods of failure recovery and analyze their complexity, like *restartable processors* in Chlebus *et al.* [7], or *stage checkpointing* in De Prisco *et al.* [8], etc. There are many specific and distinct failure recovery schemes, but they can be grouped into three broad classes:

*Under revision for *Journal of Statistical Theory and Practice.*

RESUME, also referred to as preemptive resume;
REPLACE, also referred to as preemptive repeat different;
RESTART, also referred to as preemptive repeat identical.

In the RESUME scenario, if there is a processor failure while a job is being executed, after repair is implemented the job can continue where it left off. All that is required mathematically is to remember the state of the system when failure occurred. In the REPLACE situation, if a job fails, it is replaced by a different job having the same distribution. Here, no details concerning the previous job are necessary in order to continue.

The analysis of the distribution function $H(x) = \mathbb{P}(X \le x)$ when the policy is RESUME or REPLACE was carried out by Kulkarni *et al.* [11], [12] (see also Bobbio & Trivedi [4], Castillo & Siewiorek [5] and Chimento & Trivedi [6]). The RESTART policy had resisted detailed analysis until the recent papers by Sheahan *et al.* [15], Asmussen *et al.* [1], Jelenkovic & Tan [10], where the tail asymptotics of H was found under a variety of assumptions on F and G. The setting of [10] is file transfer problems and involves an on-off model that incorporates what in the present setting corresponds to repairs. In contrast, [1] has its background in the computer science literature on failure recovery in the execution of a program on a computer.

For many systems failure is sufficiently rare to be ignored, or dealt with as an afterthought. For other systems, failure is common enough that the design choice of how to deal with it may have a significant impact on the performance of the system. One such example arises in parallel computing, where the probability of failure of a single processor in isolation may be small, but the number of processors is so large (in practice, often hundreds or thousands) that the probability that one or more processors fail cannot be neglected. The present paper studies the implications of the analysis of [1] for this situation. To formalize the set-up, assume that the job is split into M parts of lengths $S_1, \ldots, S_M$, which are executed on M parallel processors. The total times on the processors, including restarts, are denoted $X_1, \ldots, X_M$. Thus the total time for the whole job is $Z = \max_{i=1,\ldots,M} X_i$. What can then be said about the distribution of Z? For example, assume there is given a cost function of the type $a + bM + cZ_M$ where a is a set-up cost, b a cost per processor and c the cost per unit time. One would then want to choose M to minimize the expected cost $a + bM + c\mathbb{E}Z_M$ (note that one expects $\mathbb{E}Z_M$ to be a decreasing function of M).

The reason for using parallel processors will often be that the job is large. For example, the job could consist in generating R replicates of a Monte Carlo estimator for some large R. On the other hand, there may be situations where speed is an essential factor when executing a job of small or moderate size, i.e. the cost function has a large c. For example this could occur in filtering a noisy signal or in option price calculations based upon high-frequency input. This suggests considering a general triangular array situation where the total job size $T = T_M$ and hence $F = F_M$, the distribution of the job time faced by a single processor, varies with the number M of processors. We then write $S_1^{(M)}, \ldots, S_M^{(M)}$, $X_1^{(M)}, \ldots, X_M^{(M)}$,

$$Z_M \;=\; \max_{i=1,\ldots,M} X_i^{(M)} \,,$$

and $H_M(x) = \mathbb{P}(Z_M \le x)$. We will consider two scenarios:

(D) $T = T_M = t_M$ for some deterministic t_M and $S_M = s_M = t_M/M$; then F_M is the one-point distribution at s_M;

(Γ) F_M is Gamma(α_M, λ) with density

$$f_M(t) \; = \; \frac{\lambda^{\alpha_M}}{\Gamma(\alpha_M)} t^{\alpha_M - 1} \mathrm{e}^{-\lambda t} \, .$$

Further, $S_1^{(M)}, \ldots, S_M^{(M)}$ are independent. Thus the distribution of the total job size is a Gamma$(M\alpha_M, \lambda)$ distribution.

A random total job size arises in situations where the run length of the job sent to parallel processing will not be known in advance but is random. An example is Monte Carlo simulations involving random number generation by acceptance-rejection or more complicated stopping times such as cycles in regenerative simulation (see [2]). Note that since for fixed λ, the Gamma(α, λ) distributions form a convolution semi-group in α, assumption (Γ) is a natural stochastic extension of the deterministic set-up (D) (α_M corresponds to s_M). For example, in the Monte Carlo setting each replication could take a Gamma(α, β) time, and each processor would be asked to perform R_M replications. Then $\alpha_M = \beta R_M$. Of course, the Gamma case is only one among many where the total job size is infinitely divisible, and independence among subjobs is a reasonable assumption (such independence may certainly fail in some situations, but we do not consider this possibility here).

In scenario (D), we sometimes assume that $t_M = t_1 M^p$, i.e. $s_M = s_1 M^{p-1}$ for some $p \geq 0$. Here $p = 1$ could be relevant for the Monte Carlo example and $p = 0$ for the filtering example, though clearly in both situations intermediate values could also arise. The cases $p < 1$, $p = 1$ and $p > 1$ are qualitatively different since in the first $s_M \to 0$ and in the third $s_M \to \infty$ subject to (D), whereas s_M is constant when $p = 1$; analogous remarks apply to the Gamma case with the α_M taking the roles of the s_M.

We will assume throughout the paper that the failure time distribution G is independent of M and, except for Section 4, that G is exponential, with rate parameter μ.

The paper starts in Section 3 by an analysis of the case $p = 1$. This is fairly easy, because then S does not depend on M and the $X_{i,M}$ are i.i.d. random variables with a distribution not depending on M. Given the results from [1] on the tail of H, classical extreme value theory ([13]) can therefore be easily translated into a limit theorem for Z_M.

If $p \neq 1$, the $X_{i,M}$ have a distribution depending on M, so that we are beyond classical extreme value theory and have to consider a triangular array setting. This is carried out in Section 4 for $p < 1$ and Section 5 for $p > 1$. Finally, the Gamma case with $\alpha_M \to \infty$ is treated is Section 6 (the case $\alpha_M \to 0$ is non-trivial and is not included here).

2 Preliminaries

We first recall some background material from Asmussen *et al.* [1] for the RESTART setting with F independent of M. The key to the analysis in this work is the fact that given $T = t$, X is distributed as

$$t + S(t) \quad \text{where} \quad S(t) = \sum_{i=1}^{N(t)} U_i(t)\,, \tag{2.1}$$

where the $U_i(t)$ are i.i.d. distributed as U conditioned on $U \le t$, i.e.

$$\mathbb{P}\bigl(U_i(t) \le y\bigr) \;=\; \begin{cases} G(y)/G(t) & y \le t \\ 1 & y > t \end{cases},$$

and $N(t)$ is an independent geometric r.v. with success parameter $\overline{G}(t) = 1 - G(t)$, that is, $\mathbb{P}\bigl(N(t) = n\bigr) = \overline{G}(t)G(t)^n$. The following result plays a key role in [1] as well as the present paper:

Lemma 2.1. *Assume $T \equiv t_0$ and $\overline{G}(t_0) > 0$. Then*

$$\overline{H}(x) \;\sim\; C(t_0)\mathrm{e}^{\gamma(t_0)t_0}\mathrm{e}^{-\gamma(t_0)x} \tag{2.2}$$

where $\gamma(t) > 0$ is the solution of $\int_0^t \mathrm{e}^{\gamma(t)y}G(\mathrm{d}y) = 1$ and $C(t) = \overline{G}(t)/\gamma(t)B(t)$ where $B(t) = \int_0^t y\mathrm{e}^{\gamma(t)y}g(y)\,\mathrm{d}y$. Further,

$$\mathrm{e}^{-\gamma(t_0)x} \;\le\; \overline{H}(x) \;\le\; \mathrm{e}^{\gamma(t_0)t_0}\mathrm{e}^{-\gamma(t_0)x} \tag{2.3}$$

A common terminology refers to (2.2) as the *Cramér-Lundberg approximation* and to (2.3) as *Lundberg's inequality.*

We shall also use the following obvious consequence of the representation (2.1) of the conditional distribution of X given $T = t$:

$$\overline{H}(x) \;=\; \int_0^\infty \mathbb{P}\bigl(t + S(t) > x\bigr)\,F(\mathrm{d}x)\,. \tag{2.4}$$

For Scenario (Γ), the relevant result from [1] is the following (with some typos in [1] corrected here):

Lemma 2.2. *Consider Scenario* (Γ) *with $\alpha_M \equiv \alpha$ independent of M. Then*

$$\overline{H}(x) \;\sim\; C\frac{\log^{\alpha-1}x}{x^{\lambda/\mu}} \quad x \to \infty, \quad \text{where} \quad C \;=\; \frac{\Gamma\left(\frac{\lambda}{\mu}\right)}{\mu^{\frac{\lambda}{\mu}}}\frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{1}{\mu^\alpha}\,.$$

We shall also need:

Lemma 2.3. *Let K be a distribution function such that $M\overline{H}_M\bigl(a_M y + b_M\bigr) \to \log K(y)$ for all y. Then the distribution of $(Z_M - b_M)/a_M$ converges to K.*

The lemma is standard in extreme value theory when H_M is independent of M and follows from the fact that

$$\mathbb{P}\bigl((Z_M - b_M)/a_M \le y\bigr) \;=\; H_M\bigl(a_M y + b_M\bigr)^M \;=\; \bigl(1 - M\overline{H}_M\bigl(a_M y + b_M\bigr)/M\bigr)^M\,,$$

by taking logs and expanding in a Taylor series. The classical limits relevant for this paper are the Gumbel distribution with $K(y) = \mathrm{e}^{-\mathrm{e}^{-y}}$ and the Fréchet distribution with parameter $\beta > 0$ where $K(y) = \mathrm{e}^{-y^{-\beta}}$ (a Weibull limit may also occur in the classical setting, but is not relevant for RESTART because it requires bounded support). However, in a triangular setting other types of K's may occur, of which we will later see examples. A general reference on extreme value theory for triangular arrays is Valente Freitas & Hüsler [16]. However, this reference basically covers only a neighborhood of classical extreme value theory (i.e., S_M not too varying with M so that non-classical limits are not covered), and further, it requires a differentiability condition on H_M which fails at $s_M, 2s_M, \ldots$.

An important feature worth stressing is that extreme value statements deal with *typical values* of Z_M (of the form $b_M + a_M y$ in the setting of Lemma 2.3), not with tail behavior.

3 The case $p = 1$: classical extreme values

Assume that F_M does not depend on M.

Proposition 3.1. *Consider the case $s_M \equiv s_1$ in Scenario (D). Let γ denote the solution of $1 = \int_0^{t_0} \mathrm{e}^{\gamma y} \mu e^{-\mu y} \mathrm{d}y$ and set $C = \mathrm{e}^{-\mu t_0}/\gamma B$, where $B = \int_0^{t_0} y \mathrm{e}^{\gamma y} \mu \mathrm{e}^{-\mu y} \mathrm{d}y$. Then $\gamma\big(Z_M - s_1\big) - \log(MC)$ has a limiting Gumbel distribution as $M \to \infty$.*

Proof. Note that $Z_M - s_1$ is distributed as the maximum of M independent copies of $S(s_1)$ and that

$$\mathbb{P}\big(S(t_0) > x\big) \ \sim \ C\mathrm{e}^{-\gamma x}\,,\ x \to \infty,$$

by Lemma 2.1. An asymptotic exponential tail is a standard sufficient condition in extreme value theory for the random variable to be in the maximum domain of attraction of the Gumbel distribution, and the form of the normalizing constants also follows from this theory. A direct proof from Lemmas 2.1 and 2.3 is straightforward: with $a_M = 1/\gamma$, $b_M = s_1 + \log(MC)/\gamma$, one gets

$$M\overline{H}\big(a_M y + b_M\big) \ \sim \ MC\mathrm{e}^{\gamma s_1}\mathrm{e}^{-\gamma(a_M y + b_M\big)} \ = \ \mathrm{e}^{-y}\,.$$

□

The implication is that Z_M is of order $\log M/\gamma$. For example, since $-\log\log 2$ is the median in the Gumbel distribution, we obtain the approximation $s_1 - \log\log 2/\gamma + \log(MC)/\gamma$ for the median of Z_M; note that, as remarked at the end of Section 2, this is not a tail approximation but telling information about the typical values of M_N. Similarly, since the Euler constant $\varphi \approx 0.577$ is the mean of the Gumbel distribution, one obtains the approximation $\varphi/\gamma + s_1 + \log(MC)/\gamma$ for $\mathbb{E}Z_M$ (for verification of the required uniform integrability, see Pickands [14]).

Proposition 3.2. *Consider Scenario (Γ) with $\alpha_M \equiv \alpha$ independent of M, and define*

$$a_M \ = \ \frac{C^{\mu/\lambda}}{(\lambda/\mu)^{(\alpha-1)\mu/\lambda}} M^{\mu/\lambda} \log^{((\alpha-1)\mu/\lambda} M\,,$$

where C is defined in Lemma 2.2. Then Z_M/a_M has an approximate Fréchet distribution with parameter $\beta = \lambda/\mu$.

Proof. The result again follows from the standard extreme value characterization of the maximum domain of attraction of the Fréchet distribution and Lemma 2.2. Again, a direct proof from Lemma 2.3 is easy: with $b_M = 0$, one gets

$$MH(\alpha_M y) \sim \frac{1}{y^{\frac{\lambda}{\mu}}}\left(\frac{\log(\alpha_M y)}{\log(M)^{\frac{\mu}{\lambda}}}\right)^{\alpha-1} \to \frac{1}{y^{\frac{\lambda}{\mu}}},.$$

□

Again using the median as an example, the approximation for the median of Z_M becomes $a_M/\log^{1/\beta} 2$. The mean of the Fréchet distribution is finite if and only if $\beta > 1$ and then equals $\Gamma(1-1/\beta)$. This suggests the approximation $a_M\Gamma(1-1/\beta)$ for $\mathbb{E}Z_M$ when $\lambda > \mu$. Since a_M is roughly of order $M^{\mu/\lambda}$ which increases much faster than the $\log M$ occurring in Scenario (D), this shows the dramatic effect of randomness on the total job size.

4 Scenario (D) with $p < 1$

We now assume in Scenario (D) that $t_M = t_1 M^p$ for some $0 \le p < 1$ and $t_1 > 0$ so that $s_M = t_1 M^{p-1}$. We will work with the following condition on G:

$$G(x) = x^\alpha L(x) \tag{4.1}$$

with $\alpha > 0$ and L slowly varying at 0, so that $\lim_{x\to\infty} L((tx)^{-1})/L(x^{-1}) = 1$ $t > 0$. In particular, this covers a Gamma G where $L(x)$ has a limit as $x \downarrow 0$ (in the exponential set-up, $\alpha = 1$ and $L(x) \to \mu$).

We note the following consequence of (4.1):

$$\lim_M \mathbb{P}(s_M^{-1}U \le x \mid U \le s_M) = \lim_M \frac{G(s_M x)}{G(s_M)} = \lim_M x^\alpha \frac{L(s_M x)}{L(s_M)} = x^\alpha, \tag{4.2}$$

where $0 \le x \le 1$. We define $U^{(\alpha)}$ to be a random variable with distribution function

$$\mathbb{P}(U^{(\alpha)} \le x) = \begin{cases} 0 & x \le 0 \\ x^\alpha & 0 < x \le 1 \\ 1 & x > 1 \end{cases},$$

and because of (4.2) we have

$$s_M^{-1}U \le x \mid U \le s_M \xrightarrow{\mathcal{D}} U^{(\alpha)}.$$

Theorem 4.1.
I) *Assume $p \ne (k\alpha - 1)/k\alpha$ for any $k \in \mathbb{N}$. Set $p^* = \lfloor 1/(\alpha(1-p)) \rfloor$. Then*

$$t_1^{-1}M^{1-p}Z_M - 1 \xrightarrow{\mathbb{P}} p^*. \tag{4.3}$$

II) *Assume* $p = (k\alpha - 1)/k\alpha$ *for some* $k \in \mathbb{N}$, *and also that* $\lim_{x\downarrow 0} L(x) = \gamma \in (0, \infty]$ *exists. Then*

$$t_1^{-1} M^{1-p} Z_M - 1 \xrightarrow{\mathcal{D}} V\,, \tag{4.4}$$

where V *is distributed as*

$$\max_{1\le j\le N} \left(k-1, \sum_{i=1}^{k} U_i^j \right)$$

with the U_i^j *being i.i.d.* $U^{(\alpha)}$ *r.v.'s,* N *is an independent Poisson r.v. with mean* $\gamma^k t_1^{\alpha k}$ *when* $\gamma < \infty$, *and* $N = \infty$ *a.s. when* $\gamma = \infty$.

For the proof, we denote by $R_i^{(M)}$ the number of restarts of the ith processor, and let $V_k^{(M)}$ be the number of processors, with k restarts, so that

$$V_k^{(M)} = \sum_{i=1}^{M} I(R_i^{(M)} = k)$$

Let $\rho_M = G(s_M)$ and define

$$\Theta_{M,k} = \rho_M^k (1 - \rho_M).$$

We have $I(R_i^{(M)} = k) \stackrel{\mathcal{D}}{=} \mathrm{Bin}(1, \Theta_{M,k})$ and $V_k^{(M)} \stackrel{\mathcal{D}}{=} \mathrm{Bin}(M, \Theta_{M,k})$.

As a first step in the proof of Theorem 4.1, we examine the limit possibilities for $V_k^{(M)}$:

Proposition 4.1.
I) *If* $k < 1/\alpha(1-p)$, *then, setting* $\sigma_M = \sqrt{M\Theta_{M,k}(1-\Theta_{M,k})}$, *we have*

$$\frac{1}{\sigma_M}\left(V_k^{(M)} - M\Theta_{M,k} \right) \xrightarrow{\mathcal{D}} N(0,1), \qquad M \to \infty. \tag{4.5}$$

II) *If* $k = 1/\big(\alpha(1-p)\big)$, *and* $\lim_{x\downarrow 0} L(x) = \gamma \in (0, \infty]$ *exists, then*

$$V_k^{(M)} \xrightarrow{\mathcal{D}} \mathrm{Po}(\mathrm{t}_1^{\alpha\mathrm{k}} \gamma^{\mathrm{k}}) \qquad \mathrm{M} \to \infty\,, \tag{4.6}$$

where $\gamma = \infty$ *corresponds to the degenerate case at* ∞.
III) *If* $k > 1/\big(\alpha(1-p)\big)$, *then*

$$V_k^{(M)} \xrightarrow{\mathbb{P}} 0\,, \qquad M \to \infty. \tag{4.7}$$

Proof. First we notice that since $s_M = t_1 M^{p-1}$, then for all $k \in \mathbb{N}$

$$M\Theta_{M,k} = M^{1+k\alpha(p-1)} M^{-k\alpha(p-1)} G(s_M)^k \big(1 - G(s_M)\big) \tag{4.8}$$

$$\sim M^{1+k\alpha(p-1)} L(t_1 M^{p-1})^k t_1^{\alpha k} \tag{4.9}$$

Now, for the proof of I) assume $k < 1/\big(\alpha(1-p)\big)$. We need to prove that $M\Theta_{M,k} \to \infty$. This is seen by defining $H(y) = L(t_1/y)^k$. Then H is slowly varying at infinity, and we have:

$$M\Theta_{M,k} = M^{1+k\alpha(p-1)} L(t_1 M^{p-1})^k \;=\; M^{1+k\alpha(p-1)} H(M^{1-p})\,.$$

Substituting $x_M = M^{1-p}$ in this expression yields

$$x_M^{\frac{1}{1-p}-\alpha k} H(x_M)\,,$$

which tends to infinity by Proposition 1.3.6(v) in [3]. This implies that $\sigma_M \to \infty$, and therefore the normal approximation of the binomial distribution (e.g. (5.33.1) in [9]) implies

$$\frac{1}{\sigma_M} \sum_{i=1}^{M} \Big(I(R_i^{(M)} = k) - \Theta_{M,k} \Big) \xrightarrow{\mathcal{D}} N(0,1),$$

thus proving I).

The proof of III) uses the same calculation as above, where the assumption $k > 1/(\alpha(1-p))$ implies $M\Theta_{M,k} \to 0$ (again, using Proposition 1.3.6 in [3]), that is $\mathbb{E}V_k^{(M)} \to 0$, and since $V_k^{(M)} \geq 0$ we have $V_k^{(M)} \to 0$ in L^1, which proves III). Regarding II), we see that (4.6) follows from (4.9) and the Law of Small Numbers if $\gamma \in (0, \infty)$. If $\gamma = \infty$ then we may use (4.9) to conclude that $M\Theta_{M,k} \to \infty$. Using Chebycheff's inequality we have that $\mathbb{P}(V_k^M \leq \frac{M\Theta_{M,k}}{2}) \to 0$, and therefore $\lim_M \mathbb{P}(V_k^M \leq x) = 0$ for all x, which proves II). □

Corollary 4.1. *If* $k < 1/\big(\alpha(p-1)\big)$*, then* $\lim_M \mathbb{P}(V_k^{(M)} \geq x) \to 1$ *for all* x*.*

Proof. Since $M\Theta_{M,k}/\sigma_M \to \infty$

$$\lim_M \mathbb{P}(V_k^{(M)} \geq x) \;=\; \lim_M \mathbb{P}\left(\frac{V_k^{(M)} - M\Theta_{M,k}}{\sigma_M} \geq \frac{x - M\Theta_{M,k}}{\sigma_M} \right) \to 1\,.$$

□

We are now ready to prove Theorem 4.1:

Proof. In order for $t_1^{-1} M^{p-1} Z_M - 1 = s_M^{-1}(Z_M - s_M)$ to be greater than p^*, we must have at least one processor with p^*+1 restarts. Using Proposition 4.1 III), we obtain

$$\limsup_M \mathbb{P}(t_1^{-1} M^{1-p} Z_M - 1 > p^*) \;\leq\; \limsup_M \mathbb{P}(V_{p^*+1} > 0) = 0\,.$$

Let $\epsilon, \epsilon_1 > 0$ be given. We wish to show that

$$\liminf_M \mathbb{P}(t_1^{-1} M^{1-p} Z_M - 1 \geq p^* - \epsilon) \geq 1 - \epsilon_1$$

Let Z_M^* denote the random variable similar to Z_M, but where we only take the maximum over the processors with exactly p^* restarts, that is:

$$Z_M^* \;=\; \max_{1 \leq i \leq M} X_i^{(M)} I\big(R_i^{(M)} = p^*\big)$$

We see that

$$t_1^{-1} M^{1-p} Z_M^* - 1 \;\overset{\mathcal{D}}{=}\; \max_{1 \leq i \leq V_{p^*}^{(M)}} \sum_{j=1}^{p^*} t_1^{-1} M^{p-1} U_j^{(M),i}$$

where the $U_j^{(M),i}$ are independent and distributed as $s_M^{-1}U \mid U < s_M$. Since $p^* < 1/\big(\alpha(1-p)\big)$, we have by Corollary 4.1 that $V_{p^*}^{(M)} \xrightarrow{\mathbb{P}} \infty$, and therefore, for any $K \in \mathbb{N}$

$$\liminf_M \mathbb{P}\Big(\max_{1\le i\le V_{p^*}^{(M)}} \sum_{j=1}^{p^*} t_1^{-1}M^{1-p}U_j^{(M),i} \ge p^* - \epsilon \Big) \ \ge \tag{4.10}$$

$$\liminf_M \mathbb{P}\Big(\max_{1\le i\le K} \sum_{j=1}^{p^*} t_1^{-1}M^{1-p}U_j^{(M),i} \ge p^* - \epsilon \Big)\,. \tag{4.11}$$

Furthermore, since

$$\max_{1\le i\le K} \sum_{j=1}^{p^*} t_1^{-1}M^{1-p}U_j^{(M),i} \ \xrightarrow{\mathcal{D}}\ \max_{1\le i\le K} \sum_{j=1}^{p^*} U_j^i$$

where the U_j^i are i.i.d. and are distributed as U^α, we may complete the proof of I) by choosing K so large that

$$\mathbb{P}\Big(\max_{1\le i\le K} \sum_{j=1}^{p^*} U_j^i \ge p^* - \epsilon \Big) \ge 1 - \epsilon_1\,.$$

Regarding II), we see that if $k = 1/\big(\alpha(1-p)\big)$ then by (4.6) we have asymptotically N processors which have k restarts, where $N \sim \mathrm{Po}(t_1^{\alpha k}\gamma^k)$; by (4.5) we have infinitely many processors with $k-1$ restarts, and by (4.7) we have 0 processors with $k+1$ restarts. Define the following r.v.'s:

$$Z_{M,1} = \max_{1\le i\le M} X_i^{(M)} I(R_i^{(M)} < k)$$
$$Z_{M,2} = \max_{1\le i\le M} X_i^{(M)} I(R_i^{(M)} = k)$$
$$Z_{M,3} = \max_{1\le i\le M} X_i^{(M)} I(R_i^{(M)} > k)$$

Then clearly $Z_M = \max(Z_{M,1}, Z_{M,2}, Z_{M,3})$ and since $t_1^{-1}M^{1-p}Z_{M,1} - 1 \xrightarrow{\mathbb{P}} k-1$, $t_1^{-1}M^{1-p}Z_{M,3} - 1 \xrightarrow{\mathbb{P}} 0$ and $t_1^{-1}M^{1-p}Z_{M,2} - 1 \xrightarrow{\mathcal{D}} \sum_{i=1}^k U_i^j$, where (U_j^j) is an i.i.d. sequence of r.v.'s distributed as $U^{(\alpha)}$, the proof is complete. □

5 Scenario (D) with $s_M \to \infty$

We now consider Scenario (D) with $s_M \to \infty$ (for example, $t = t_M = t_1M^p$ with $p > 1$; equivalently, M grows with t like $t^{1/p}$, i.e. at a rate somewhat slower than t). That is, there is significant but not massive parallelization. Let $\gamma_M = \gamma(s_M)$ in the notation of Lemma 2.1. We shall prove:

Theorem 5.1. *Consider Scenario (D) with $s_M \to \infty$. Then $\mu e^{-\mu s_M} Z_M - \log M$ has a limiting Gumbel distribution as $M \to \infty$.*

This means that Z_M is of order $\mathrm{e}^{\mu s_M} \log M/\mu = \mathrm{e}^{\mu t_M/M} \log M/\mu$.

Lemma 5.1. *Let $\gamma_M = \gamma(s_M)$ in the notation of Lemma 2.1. Then $\gamma_M - \mu\mathrm{e}^{-\mu s_M} = \mathrm{O}\big(s_M \mathrm{e}^{-2\mu s_M}\big)$.*

Proof. Evaluating the integral in the defining equation

$$1 \;=\; \int_0^{s_M} \mathrm{e}^{\gamma_M y} \mu \mathrm{e}^{-\mu y}\,\mathrm{d}y$$

explicitly, one gets

$$1 \;=\; \frac{\mu}{\mu - \gamma_M}\big(1 - \mathrm{e}^{-(\mu-\gamma_M)s_M}\big)\,,$$

which can be rewritten as

$$\gamma_M \;=\; \mu \mathrm{e}^{-(\mu-\gamma_M)s_M}\,. \tag{5.1}$$

This shows that γ_M is of first order $\mu\mathrm{e}^{-\mu s_M}$ (as is shown already in [1]). In particular, $\gamma_M s_M \to 0$ so that by Taylor expansion of (5.1),

$$\gamma_M \;\approx\; \mu\mathrm{e}^{-\mu s_M}(1 + \gamma_M s_M)\,.$$

This proves the assertion. □

Proof of Theorem 5.1. Let F_M denote the distribution of $X_i^{(M)}$. Then by Lundberg's inequality,

$$\mathrm{e}^{-\gamma_M x} \;\le\; \overline{H}_M(x) \;\le\; \mathrm{e}^{\gamma_M s_M}\mathrm{e}^{-\gamma_M x}\,.$$

Let $b_M = \log M/\gamma_M$, $a_M = 1/\gamma_M$. Then

$$M\overline{H}_M(a_M y + b_M) \;\ge\; M\mathrm{e}^{-\gamma_M(a_M y + b_M)} \;=\; M\mathrm{e}^{-y+\log M} \;=\; \mathrm{e}^{-y}\,.$$

Similarly,

$$M\overline{H}_M(a_M y + b_M) \;\le\; M\mathrm{e}^{\gamma_M s_M}\mathrm{e}^{-\gamma_M(a_M y + b_M)} \;\to\; 1 \cdot \mathrm{e}^{-y}\,.$$

Thus $M\overline{H}_M(a_M y + b_M) \;\to\; \mathrm{e}^{-y}$ for all y, which implies that

$$\gamma_M Z_M - \log M \;=\; \frac{Z_M - b_M}{a_M}$$

has a Gumbel limit.

It then follows that Z_M is roughly of order $1/\gamma_M$ or equivalently $\mathrm{e}^{\mu s_M}$. To replace γ_M by $\mu\mathrm{e}^{-\mu s_M}$ in the limit statement for Z_M, one therefore needs

$$(\gamma_M - \mu\mathrm{e}^{-\mu s_M})\mathrm{e}^{\mu s_M} \;\to\; 0\,,$$

which follows by Lemma 5.1. □

6 The Gamma case

We now consider Scenario (Γ) with $\alpha_M \to \infty$.

Theorem 6.1. *Consider the Gamma case with* $\alpha_M \to \infty$ *and let* $r = \mu/\lambda$. *Assume in addition that* $\alpha_M/\log M \to \infty$. *Then* Z_M *is of logarithmic order* $\mathrm{e}^{r\alpha_M}$ *in the sense that* $\log Z_M/\alpha_M \xrightarrow{\mathbb{P}} r$ *as* $M \to \infty$.

For the proof, define $x_M = \mathrm{e}^{r_1\alpha_M}$. We shall show that

$$M\overline{H}_M(x_M) \;\to\; \begin{cases} \infty & \text{if } r_1 < r \\ 0 & \text{if } r_1 > r. \end{cases} \tag{6.1}$$

Indeed, if $r_1 < r$ then (6.1) shows that the expected number of processors i with $X_{i,M} > x_M$ tends to ∞ and hence the probability that one $X_{i,M} > x_M$ tends to 1. Similarly, if $r_1 > r$ then (6.1) shows that the expected number of processors i with $X_{i,M} > x_M$ tends to 0, and hence so does the probability that one $X_{i,M} > x_M$.

Lemma 6.1. *Define*

$$I_M \;=\; \int_0^{cx_M} a^{\lambda/\mu-1}\mathrm{e}^{-a}\varphi_M(a)\,\mathrm{d}a \;=\; \int_0^{cx_M} a^{1/r-1}\mathrm{e}^{-a}\varphi_M(a)\,\mathrm{d}a$$

where $0 < c \le \mu'$ *is a constant and*

$$\varphi_M(a) \;=\; \Big(1 + \frac{\log\mu' - \log a}{\log x_M}\Big)^{\alpha_M-1}.$$

Then $I_M \to {\mu'}^{1/r_1}\Gamma(1/r - 1/r_1)$ *as* $M \to \infty$ *when* $r_1 > r$, *whereas*

$$\liminf_{M\to\infty} \frac{I_M}{\alpha_M^{1/2}\exp\big\{(\delta - \log\delta - 1)\alpha_M\big\}} \;>\; 0$$

when $r_1 < r$, *where* $\delta = r_1/r$.

Note that the convexity of the log implies that $\delta - \log\delta - 1 > 0$ when $\delta \ne 1$.

Proof. We split I_M up into the contributions I'_M and I''_M from $a < \mu'$ and $\mu' < a < cx_M$, respectively. For $a < \mu'$, $\varphi_M(a) \uparrow {\mu'}^{1/r_1}a^{-1/r_1}$, and hence by monotone convergence,

$$I'_M \;\uparrow\; \int_0^{\mu'} a^{1/r-1/r_1-1}\mathrm{e}^{-a}\,\mathrm{d}a\,.$$

When $r_1 > r$, we thus need in addition to show that

$$I''_M \;\to\; {\mu'}^{1/r_1}\int_{\mu'}^{\infty} a^{1/r-1/r_1-1}\mathrm{e}^{-a}\,\mathrm{d}a\,.$$

This follows by dominated convergence since $\varphi_M(a)$ is dominated by 1 on (μ',∞) and has the limit ${\mu'}^{1/r_1}a^{-1/r_1}$.

Consider now the case $r_1 < r$. Substituting

$$y \;=\; 1 + (\log \mu' - \log a)/\log x_M \;=\; 1 + (\log \mu' - \log a)/r_1\alpha_M\,,$$

we have

$$\log a = \log \mu' + (1-y)r_1\alpha_M, \quad \frac{1}{a}\mathrm{d}a = -r_1\alpha_M\,\mathrm{d}y, \quad a = \mu \mathrm{e}^{r_1\alpha_M}\mathrm{e}^{-r_1\alpha_M y}\,.$$

Thus, bounding e^{-a} below by $c_1 = \mathrm{e}^{-\mu}$, we get

$$\begin{aligned} I_M \;\ge\; I'_M \;&\ge\; c_2\alpha_M \mathrm{e}^{\delta\alpha_M}\int_1^\infty y^{\alpha_M-1}\mathrm{e}^{-y\delta\alpha_M}\mathrm{d}y \\ &=\; c_2\mathrm{e}^{\delta\alpha_M}\delta^{-\alpha_M}\alpha_M^{1-\alpha_M}\int_{\delta\alpha_M}^\infty z^{\alpha_M-1}\mathrm{e}^{-z}\,\mathrm{d}z\,. \end{aligned}$$

The last integral divided by $\Gamma(\alpha_M)$ is the probability that a Gamma$(\alpha_M, 1)$ r.v. exceeds $\delta\alpha_M$. Since this probability goes to 1 when $\delta < 1$, we get

$$I_M \;\ge\; c_3\mathrm{e}^{\delta\alpha_M}\delta^{-\alpha_M}\alpha_M^{1-\alpha_M}\Gamma(\alpha_M)\,.$$

Using Stirling's approximation

$$\Gamma(\alpha_M) \;\sim\; \mathrm{e}^{-\alpha_M}\alpha_M^{\alpha_M-1}\sqrt{2\pi\alpha_M}$$

completes the proof. □

Proof of (6.1). Let first $r_1 < r$. By the Lundberg lower bound, we have for any $\mu' > \mu$ and some t_0 that

$$\overline{H}_M(x_M) \;\ge\; \int_{t_0}^\infty \mathrm{e}^{-\mu'\mathrm{e}^{-\mu t}x_M}\frac{\lambda^{\alpha_M}}{\Gamma(\alpha_M)}t^{\alpha_M-1}\mathrm{e}^{-\lambda t}\,\mathrm{d}t\,. \tag{6.2}$$

Substituting $a = \mu'\mathrm{e}^{-\mu t}x_M$, we have

$$t \;=\; \frac{1}{\mu}(\log\mu + \log x_M - \log a)\,, \quad \mathrm{d}t \;=\; -\frac{1}{\mu a}\,\mathrm{d}a\,,$$

and thus (6.2) becomes

$$\begin{aligned} &\frac{\lambda^{\alpha_M}}{\Gamma(\alpha_M)\mu^{\alpha_M}x_M^{\lambda/\mu}}\int_0^{\mu'\mathrm{e}^{-\mu t_0}x_M} a^{\lambda/\mu-1}\mathrm{e}^{-a}\big(\log\mu' + \log x_M - \log a\big)^{\alpha_M-1}\,\mathrm{d}a \\ &= \frac{1}{\Gamma(\alpha_M)r^{\alpha_M}\mu'^{\mu/\lambda}}\frac{\log^{\alpha_M-1}x_M}{x_M^{1/r}}\int_0^{\mu'\mathrm{e}^{-\mu t_0}x_M} a^{\lambda/\mu-1}\mathrm{e}^{-a}\varphi_M(a)\,\mathrm{d}a\,. \end{aligned}$$

By Lemma 6.1, this implies that $M\overline{H}_M(x_M)$ is of larger order than

$$M\frac{1}{\Gamma(\alpha_M)r^{\alpha_M}}\frac{\log^{\alpha_M-1}x_M}{x_M^{1/r}}\alpha_M^{1/2}\exp\bigl\{(\delta-\log\delta-1)\alpha_M\bigr\}\,.$$

By Stirling's approximation, this is in turn of order

$$\begin{aligned}
&\frac{M\mathrm{e}^{\alpha_M}}{\alpha_M^{1/2}\alpha_M^{\alpha_M-1}r^{\alpha_M}}\frac{\log^{\alpha_M-1}x_M}{x_M^{1/r}}\\
&\quad= \frac{M\mathrm{e}^{\alpha_M}}{\alpha_M^{1/2}\alpha_M^{\alpha_M-1}r^{\alpha_M}}\frac{\alpha_M^{\alpha_M-1}r_1^{\alpha_M-1}\alpha_M^{\alpha_M-1}}{\mathrm{e}^{(r_1/r)\alpha_M}}\alpha_M^{1/2}\exp\bigl\{(\delta-\log\delta-1)\alpha_M\bigr\}\\
&\quad= M\exp\bigl\{(1-\delta+\log\delta)\alpha_M\bigr\}/\alpha_M^{1/2}\cdot\alpha_M^{1/2}\exp\bigl\{(\delta-\log\delta-1)\alpha_M\bigr\}\\
&\quad= M \;\to\; \infty\,.
\end{aligned}$$

Now let $r_1 > r$. Choose μ' such that $\mu'\mathrm{e}^{-\mu t} \le \gamma(t)$ for $t \ge 1$ and let $c_4 = \mu'\sup_{t\ge 1} t\mathrm{e}^{-\mu t} = \mathrm{e}^{-1}$. Then by the upper Lundberg bound,

$$\overline{H}(x_M) \;\le\; \mathbb{P}(T_M\le 1) \;+\; c_4\int_1^\infty \mathrm{e}^{-\mu'\mathrm{e}^{-\mu t}x_M}\frac{\lambda^{\alpha_M}}{\Gamma(\alpha_M)}t^{\alpha_M-1}\mathrm{e}^{-\lambda t}\,\mathrm{d}t$$

Here $\mathbb{P}(T_M \le 1)$ goes to 0 at least exponentially fast in α_M. Using the same substitution as when $r_1 < r$, the integral becomes

$$\begin{aligned}
&\frac{\lambda^{\alpha_M}}{\Gamma(\alpha_M)\mu^{\alpha_M}x_M^{\lambda/\mu}}\int_0^{\mu'\mathrm{e}^{-\mu}x_M} a^{\lambda/\mu-1}\mathrm{e}^{-a}\bigl(\log\mu'+\log x_M-\log a\bigr)^{\alpha_M-1}\,\mathrm{d}a\\
&\quad= \frac{1}{\Gamma(\alpha_M)r^{\alpha_M}\mu'^{\mu/\lambda}}\frac{\log^{\alpha_M-1}x_M}{x_M^{1/r}}\int_0^{\mu'\mathrm{e}^{-\mu}x_M} a^{\lambda/\mu-1}\mathrm{e}^{-a}\varphi_M(a)\,\mathrm{d}a. \qquad (6.3)
\end{aligned}$$

Here the last integral is $\mathrm{O}(1)$ by Lemma 6.1, and using Stirling's approximation as above shows that (6.3) is of order

$$M\exp\bigl\{(1-\delta+\log\delta)\alpha_M\bigr\}/\alpha_M^{1/2}\,.$$

Putting these estimates together, recalling that $\alpha_M/\log M \to \infty$ and that $1-\delta+\log\delta < 0$ for all $\delta\ne 1$ we see that $M\overline{H}_M(x_M)\to 0$. □

References

[1] S. Asmussen, P. Fiorini, L. Lipsky, T. Rolski & R. Sheahan (2007) On the distribution of total task times for tasks that must restart from the beginning if failure occurs. *Submitted.* Available from `www.thiele.au.dk`.

[2] S. Asmussen & P.W. Glynn (2007) *Stochastic Simulation: Algorithms and Analysis.* Springer-Verlag

[3] N.H. Bingham, C.M. Goldie, & J.L. Teugels (1987) *Regular Variation.* Cambridge University Press.

[4] A. Bobbio & K. Trivedi (1990) Computation of the distribution of the completion time when the work requirement is a PH random variable. *Stochastic Models* **6**, 133–150.

[5] X. Castillo & D.P. Siewiorek (1980) A performance-reliability model for computing systems. *Proc FTCS-10, Silver Spring, MD, IEEE Computer Soc.*, 187–192.

[6] P.F. Chimento, Jr. & K.S. Trivedi (1993) The completion time of programs on processors subject to failure and repair. *IEEE Trans. on Computers* **42**(1).

[7] B.S. Chlebus, R. De Prisco, & A.A. Shvartsman (2001) Performing tasks on synchronous restartable message-passing processors. *Distributed Computing* **14**, 49–64.

[8] R. DePrisco, A. Mayer, & M. Yung (1994) Time-optimal message-efficient work performance in the presence of faults. *Proc. 13th ACM PODC*, 161–172.

[9] J. Hoffmann-Jørgensen (1994) *Probability With a View Yoward Statistics, Vol.* **I**. Chapman & Hall.

[10] P. Jelenkovic & V. Tan (2007) Can retransmission of superexponential documents cause subexponential delays? *Proc. IEEE Infocom2007*, pp. 892–900, Anchorage, 6–12 May 2007.

[11] V. Kulkarni, V. Nicola, & K. Trivedi (1986) On modeling the performance and reliability of multimode systems. *The Journal of Systems and Software* **6**, 175–183.

[12] V. Kulkarni, V. Nicola, & K. Trivedi (1987) The completion time of a job on a multimode system. *Adv. Appl. Probab.* **19**, 932–954.

[13] M.R. Leadbetter, G. Lindgren & H. Rootzén (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag.

[14] J. Pickands, III (1968) Moment convergence of sample extremes. *Ann. Math. Statist.* **39**, 881–889.

[15] R. Sheahan, L. Lipsky, P. Fiorini & S. Asmussen (2006) On the distribution of task completion times for tasks that must restart from the beginning if failure occurs. *ACM SIGMETRICS Performance Evaluation Review* **34**, 24–26.

[16] A. Valente Freitas & J. Hüsler (2003) Conditions for the convergence of maxima of random triangular arrays. *Extremes* **6**, 381–394.