Reservation and Checkpointing Strategies for Stochastic Jobs

Ana Gainaru*, Brice Goglin†, Valentin Honoré†, Guillaume Pallez (Aupy)†, Padma Raghavan*, Yves Robert†, Hongyang Sun*
*Department of EECS, Vanderbilt University, Nashville, TN, USA
†Inria, LaBRI, Univ. Bordeaux, Talence, France
‡Laboratoire LIP, ENS Lyon, France & University of Tennessee Knoxville, USA

Abstract—In this paper, we are interested in scheduling and checkpointing stochastic jobs on a reservation-based platform, whose cost depends both (i) on the reservation made, and (ii) on the actual execution time of the job. Stochastic jobs are jobs whose execution time cannot be determined easily. They arise from the heterogeneous, dynamic and data-intensive requirements of new emerging fields such as neuroscience. In this study, we assume that jobs can be interrupted at any time to take a checkpoint, and that job execution times follow a known probability distribution. Based on past experience, the user has to determine a sequence of fixed-length reservation requests, and to decide whether the state of the execution should be checkpointed at the end of each request. The objective is to minimize the expected cost of a successful execution of the jobs. We provide an optimal strategy for discrete probability distributions of job execution times, and we design fully polynomial-time approximation strategies for continuous distributions with bounded support. These strategies are then experimentally evaluated and compared to standard approaches such as periodic-length reservations and simple checkpointing strategies (either checkpoint all reservations, or none). The impact of an imprecise knowledge of checkpoint and restart costs is also assessed experimentally.

Index Terms—scheduling, checkpointing, stochastic job, reservation-based platform, reservation sequence, neuroscience application.

I. INTRODUCTION

In this paper, we revisit our recent work on reservation strategies for stochastic jobs [3]. Stochastic jobs originate from Big Data or Machine Learning workloads, whose performance is widely dependent on characteristics of input data. Figure 1 shows an example of a Neuroscience job. Reservation strategies provide a sequence of fixed length reservations to execute a stochastic job. If the reservation is too short for the job, it is restarted in a longer reservation. We extend the approach to include the possibility of checkpointing at the end of some (well-chosen) reservations. The idea of checkpointing is very natural and widely used in practice, in particular for long jobs lasting several hours, but it dramatically complicates the design of scheduling strategies. To the best of our knowledge, existing approaches either checkpoint at the end of all reservations, or never. For large-scale applications, checkpointing to save intermediate results at the end of each reservation is the de facto standard approach.

We use an example to help understand the challenges of the problem under study. Consider the jobs depicted in Figure 1. We model their execution time with $D$, a truncated LogNormal probability distribution on the domain $[a, b] = [0, 80h]$ (mean $\mu = 21h$, standard deviation $\sigma = 20h$). The exact execution time $X$ of the next job to be scheduled is not known until that job has successfully completed, but instead is randomly and uniformly sampled from the target probability distribution $D$. We want to minimize the expected cost of scheduling this job. To do so, we have to derive a sequence of reservations. Then we compute the cost of the job given that sequence, and aim at minimizing the expected value. To determine the cost of a reservation, we use the generic model from our previous work [3]. This model has been shown to encompass a variety of scenarios, ranging from the Reserved Instances of Cloud Computing where one pays (for a cheaper cost) only the reserved time [2], to High-Performance Computing (HPC) platforms where one pays the total execution time (wait time and runtime).

Specifically, for a reservation of length $W_1$ and an actual execution duration of length $X$, the cost is expressed as:

$$\alpha W_1 + \beta \min(W_1, X) + \gamma$$

(1)

where $\alpha$, $\beta$ and $\gamma$ are constant parameters that depend on the platform and the cost model. The first component $\alpha W_1$ is proportional to the reservation length (pay for what you ask). The second component $\beta \min(W_1, X)$ is proportional to the actual execution time (pay for what you use). Finally, the third and last component is a start-up time possibly associated with the first and/or second components.

To illustrate the contribution of this work, we use $\alpha = 1, \beta = \gamma = 0$ in the example, and we divide execution costs...
by a factor 60 for simplicity, so that \([a, b] = [0, 80]\) minutes. In Figure 2, we depict three strategies, and their expected costs: (i) \(S_1\) (Standard), which reserves the upper bound of \(D\), \(W_1 = b = 80\); (ii) \(S_2\) (No Checkpoint), which introduces a first reservation of size \(W_1 = 20\) before the second reservation \(W_2 = 80\); (iii) \(S_3\) (With Checkpoint), which introduces a first checkpointed reservation of size \(W_1 = 20 + 7\) (20 to cover jobs shorter than 20, and 7 (red box) is the cost to checkpoint), then a second non-checkpointed reservation of size \(W_2 = 7 + 20\) (7 (green box) is the cost to restart, 20 to cover jobs larger than 20 and smaller than 40), and a third reservation of size \(W_3 = 7 + 60\) (7 is the cost to restart, 60 to cover jobs of size up to \(b\)). Here are the expected costs of these strategies:

\[
\begin{align*}
\mathbb{E}(S_1) &= 80 \\
\mathbb{E}(S_2) &= 20 \cdot \mathbb{P}(X \leq 20) + 100 \cdot \mathbb{P}(20 < X \leq 80) \\
&= 20 \cdot 0.66 + 100 \cdot 0.34 = 47.2 \\
\mathbb{E}(S_3) &= 27 \cdot \mathbb{P}(X \leq 20) + 54 \cdot \mathbb{P}(20 < X \leq 40) + 121 \cdot \mathbb{P}(40 < X) \\
&= 27 \cdot 0.66 + 54 \cdot 0.26 + 121 \cdot 0.08 = 41.54
\end{align*}
\]

Note that \(S_3\), the variant of \(S_3\) where the second reservation is also checkpointed, would have a larger expected cost due to this second checkpoint: \(\mathbb{E}(S_3) = 27 \cdot 0.66 + 61 \cdot 0.26 + 128 \cdot 0.08 = 43.92\). Similarly one can verify that not performing the second reservation at all would have also increased the expected cost. This example shows that checkpointing does help for some scenarios but has too much overhead for others, and suggests that finding the best trade-off is difficult.

Indeed, in the general case, one has to decide which reservations should be checkpointed, depending on application profile and platform parameters. Moreover, determining the expected cost of a given reservation sequence together with scheduling decisions gets quite complicated. Section \(\text{II-A}\) gives a detailed formula for the expected cost, and Theorem \(\text{II-A}\) in Section \(\text{II-A}\) provides a simplified version. In our previous work without checkpoints \(\text{[5]}\), we have been able to analytically characterize the optimal sequence of reservations for any smooth probability distribution (except the length of the first reservation which had to be found numerically). The problem with checkpoints is dramatically more difficult, but we provide a holistic approach: we show how to compute the optimal solution for any discrete probability distribution, using a sophisticated dynamic programming algorithm. Then we show how to approximate the optimal solution for any continuous probability distribution with bounded support, by providing a reservation sequence (and its checkpointing decisions) whose expected cost is arbitrarily close to the optimal one. In practice, the restriction to bounded support is not a limitation. Given, say, a LogNormal or Weibull probability distribution defined on \([a, \infty)\), it is very natural to truncate it on a bounded interval \([a, b]\) where \(b\) corresponds to the quantile \(Q(1 - \varepsilon)\) for a small value of \(\varepsilon\). This amount to discarding job execution times that are unreasonably too long, and never encountered in practice.

The main contributions of this work are the following:

- The characterization of an optimal reservation sequence, together with its checkpointing decisions, for any discrete probability distribution, using a sophisticated dynamic programming algorithm.
- An approximation of the optimal solution for any continuous probability distribution with bounded support, by providing an algorithm to compute a reservation sequence (and its checkpointing decisions) whose expected cost is arbitrarily close to the optimal one.
- An extensive set of simulation results as well as experiments on a multicore platform, using nine probability distributions and neuroscience application traces, showing the efficiency of our strategies in a HPC environment.

The rest of the paper is organized as follows. Section \(\text{II}\) introduces the framework and main notations, and provides a detailed formula for the expected cost of a reservation sequence and its checkpointing decisions. Section \(\text{III}\) describes our key algorithmic contributions. Section \(\text{IV}\) is devoted to experimental evaluation and comparison with existing approaches. Section \(\text{V}\) evaluates the actual performance of the approach for a neuroscience application on an HPC platform. Section \(\text{VI}\) presents related work. Finally, we provide concluding remarks and hints for future work in Section \(\text{VII}\).

II. FRAMEWORK

In this section, we introduce some notations and formally define the optimization problem under study.

A. Stochastic jobs

We consider stochastic jobs whose execution times are unknown but (i) deterministic with respect to input data, so that two successive executions of the same job will have the same duration; and (ii) randomly and uniformly sampled from a given probability distribution law \(D\), whose density function (PDF) is \(f\) and cumulative distribution function (CDF) is \(F\). The probability distribution is assumed to be nonnegative, since we model execution times, and it is defined either on a finite support \([a, b]\), where \(0 \leq a < b\), or on an infinite support \([a, \infty)\) where \(a \geq 0\). Hence, the execution time of a job is a random variable \(X\), and \(\mathbb{P}(X \leq T) = F(T) = \int_{a}^{T} f(t)dt\). For notational convenience, we sometimes extend the domain of \(f\) outside the support of \(D\) by letting \(f(t) = 0\) for \(t \in [0, a] \cup [b, \infty)\).

In addition, we assume that we can interrupt the jobs at any time (divisible load application) to take a checkpoint: this will save the current progress of the execution, and enable
There is a restart if and only if there has been a checkpoint 

testion, and a sequence of reservations

restart from the last checkpoint (or from the very beginning if there was no checkpoint taken during the last reservation, the work done during that reservation is lost, and the execution must restart from the last checkpoint (or from the very beginning if no checkpoint was taken yet).

Altogether, the user needs to schedule a (possibly infinite) sequence of reservations \( W = (W_1, W_2, \ldots, W_i, W_{i+1}, \ldots) \) to execute any job whose execution time follows the distribution \( D \), and to launch these reservations one after the other, until the job successfully terminates within the duration of some reservation. In addition, the user should decide whether to take a checkpoint or not at the end of each reservation.

**Definition 1** (Reservation sequence for \( D \)). Given a probability distribution \( D \), a reservation sequence \( S = \{(W_1, δ_1), (W_2, δ_2), \ldots\} \), is defined as a sequence of reservation lengths \( W_k \) and a sequence of checkpointing decisions \( δ_k \in \{0, 1\} \): \( δ_k = 1 \) means the \( k \)-th reservation ends with a checkpoint, and \( δ_k = 0 \) means it does not.

Then, the \( k \)-th reservation can be decomposed into:

\[
W_k = R_k + T_k + C_k
\]  

where \( R_k \) is the time spent for restart, \( T_k \) for actual job execution, and \( C_k \) for checkpoint. We have \( C_k = δ_k C \) by definition. There is a restart if and only if there has been a checkpoint at some point before, hence \( R_k = (1 - \prod_{i=1}^{k-1}(1 - δ_i))R \) (assuming \( R_1 = 0 \) for the first reservation). But it is hard to keep track of actual job progress when using only the \( \{W_k, δ_k\} \) values. Consider for instance the following sequence \( S = \{W_1, (W_2, 0), (W_3, 1), (W_4, 0)\} \), which is depicted in Figure 3. If the actual job duration is \( X = t \), during which reservation will the job complete its execution? We introduce another view of the reservation sequence \( S \) by introducing the milestones \( \{t_k\} \) as shown in Figure 4. A milestone \( t_k \) represents the amount of work that has been actually executed at the end of the \( k \)-th reservation. Then, the last reservation for the job of length \( t \) is \( W_k \), where \( t_{k-1} \leq t \leq t_k \). Of course, we need that \( t_k \leq t_4 \) for all values of \( D \) (equivalently, the upper bound of the support of \( D \) is \( t \leq t_4 \) for all jobs to complete successfully with the four reservations of \( S \).

The relationship between the milestone \( t_k \) (actual work progress) and the value of \( T_k \) (time spent computing during reservation \( W_k \); see Equation (2)) is \( t_k = T_k + \sum_{i=1}^{k-1} δ_i T_i \); Indeed, the work actually progresses only from the last checkpoint, while the work executed during the previous non-checkpointed reservations is lost whenever these non-checkpointed reservations do not allow for the full completion of the job. Another way to express the relationship between \( t_k \) and \( T_k \) is the following:

\[
t_k = T_k + \max\{|t_i | 1 \leq i \leq k - 1 \text{ and } δ_i = 1\} \tag{3}
\]

Equation (3) gives a recursive way to compute \( t_k \) from its definition. We recapitulate the relations between all notations introduced in Figures 3 and 4:

\[
W_k = R_k + T_k + C_k; \quad R_k = (1 - \prod_{i<k}(1 - δ_i))R; \quad T_k = t_k - \sum_{i<k} δ_i T_i; \quad C_k = δ_k C.
\]  

In the following, we use milestones \( t_k \) rather than reservation lengths \( W_k \) to characterize a reservation sequence, and we write \( S = \{(t_1, δ_1), (t_2, δ_2), \ldots\} \) instead of \( S = \{(W_1, δ_1), (W_2, δ_2), \ldots\} \), because it is easier to use milestones when computing the expected cost of a sequence, as shown below. For notational convenience, we define \( t_0 = 0 \) as the first milestone of each sequence \( S \). Note also that we can restrict to sequences where \( t_{k-1} < t_k \), because otherwise (if \( t_{k-1} = t_k \)), the execution does not progress during the \( k \)-th reservation.

**C. Expected cost**

Given a reservation sequence \( S = ((t_i, δ_i)) \), and a job with execution time \( t \) such that \( t_{k-1} < t \leq t_k \), the cost of the sequence for that job is given by:

\[
C_S(k, t) = \sum_{i=1}^{k-1} (αW_i + βW_i + γ) + αW_k + β(R_k + t - (t_k - T_k)) + γ
\]  

where the first part is the total cost from the \( k - 1 \) first reservations that did not allow the job to complete, and the second part is the cost of the \( k \)-th reservation. The actual execution time during the \( k \)-th reservation is \( t - (t_k - T_k) \).
because $t_k - T_k$ is the amount of work done up to the beginning of that reservation; we add the restart time ($R_k$) but do not need to checkpoint (if $\delta_k = 1$) because the job successfully completes before it is taken.

We let $k(t) = k$ for a job of length $t$ such that $t_{k-1} < t \leq t_k$. Now, given a random variable $X$ following a distribution $\mathcal{D}$, the expected cost of the reservation sequence $S$ is

$$E(S(X)) = \int_0^\infty C_S(k(t), t) f(t) dt = \sum_{k=1}^{\infty} \sum_{t_{k-1}}^{t_k} C_S(k(t), t) f(t) dt \quad (6)$$

D. Optimization problem

We are now ready to state the optimization problem:

**Definition 2 (STOCHASTIC).** Given a random variable $X$ (with PDF $f$ and CDF $F$) for the execution times of a stochastic job, and a cost function in Equation (5) (with parameters $\alpha > 0$ and $\beta, \gamma \geq 0$), find a reservation strategy $S$ with minimal expected cost $E(S(X))$ as given in Equation (6).

We further define RESERVATIONONLY to be the instance of STOCHASTIC where the cost is a linear function of the reservation length only, i.e., when $\beta = \gamma = 0$. For RESERVATIONONLY, we can further consider $\alpha = 1$ without loss of generality. For instance, such costs are incurred when making reservations of resources to schedule jobs on some cloud platforms, with hourly or daily rates. Throughout the paper, we focus on the usual probability distributions, hence we assume that the density function $f$ and the CDF $F$ of $\mathcal{D}$ are smooth (infinitely differentiable), and that $\mathcal{D}$ has finite expectation.

### III. Algorithms

In this section, we establish some key properties of an optimal solution in the general setting.

A. Expected cost

We start by establishing a simpler expression for the expected cost function of STOCHASTIC in the following theorem. The proof is omitted due to space constraint but can be found in the companion report [9].

**Theorem 1.** Given a random variable (RV) $X$ and a reservation sequence $S = ((t_1, \delta_1), (t_2, \delta_2), \ldots)$, the expected cost $E(S(X))$ of a strategy $S$ given by Equation (6), with parameters $\alpha, \beta, \gamma$, can be rewritten as

$$E(S(X)) = \beta \cdot E[X] + \alpha (t_1 + \delta_1 C) + \gamma + \sum_{i=2}^{\infty} \left( \alpha W_i + \beta (R_i + (1 - \delta_{i-1}) T_{i-1} + C_{i-1}) + \gamma \right) \cdot P(X > t_{i-1}) \quad (7)$$

For ease of reading, when there is no ambiguity on the RV $X$, we write $E(S(X)) = E(S)$.

B. Dynamic programming for discrete distributions

We study the problem for a finite discrete distribution: $Y \sim (v_i, f_i)_{1 \leq i \leq n}$, where $v_i < v_{i+1}$ for all $1 \leq i \leq n - 1$ and $f_i = P(Y = v_i)$. We assume that $f_n \neq 0$ and $\sum_{i=1}^{n} f_i = 1$. Consider a strategy $S = \{(t_1, \delta_1), (t_2, \delta_2), \ldots, (t_{|S|}, \delta_{|S|})\}$, where $t_i = v_{\pi(i)}$ and $t_i < t_{i+1}$ for all $1 \leq i \leq |S| - 1$. Also, the last reservation is necessarily $t_{|S|} = v_n$ to ensure that the expected cost of the strategy is finite. By convention, we let $t_0 = v_0 = a$, hence $P(Y > t_0) = 1$. Note that we can safely restrict to strategies where each milestone $t_i$ is equal to some threshold $v_j$ of the discrete distribution: otherwise, replacing $t_i$ by the largest $v_j$ such that $v_j < t_i$ leads to a smaller cost.

Rewriting Equation (7) with $W_i = R_i + T_i + C_i$, and since $W_0 = 0$, the expected cost of strategy $S$ can be expressed as:

$$E(S) = \beta \cdot E[Y] + \sum_{i=1}^{\infty} \left( (\alpha (R_i + T_i + C_i) + \beta R_i + \gamma) \cdot P(Y > t_{i-1}) \right.$$

$$+ \sum_{i=1}^{\infty} \beta ((1 - \delta_i) T_i + C_i) \cdot P(Y > t_i) \quad (8)$$

Based on Equation (8), and using Equations (4), we construct a dynamic programming algorithm to compute the optimal reservation sequence:

**Theorem 2.** For a discrete distribution $Y \sim (v_i, f_i)_{1 \leq i \leq n}$, the optimal expected cost is returned by $E_{\text{ckpt}}(0, 0)$, where, for $0 \leq i_c \leq i_i \leq n$, $E_{\text{ckpt}}(i_c, i_i)$ is:

$$= \beta \cdot E[Y], \quad \text{if } i_i = n$$

$$= \min_{i_c + 1 \leq i_j \leq n, \Delta_j \in \{0, 1\}} \left( E_{\text{ckpt}}(\Delta_j, j) + (\alpha (v_j + \Delta_j C) + \gamma) \cdot \sum_{k=i_j+1}^{n} f_k \right.$$

$$+ \beta ((1 - \Delta_j) v_j + \Delta_j C) \cdot \sum_{k=i_j+1}^{n} f_k \left. \right), \quad \text{if } i_c = 0$$

$$= \min_{i_c + 1 \leq i_j \leq n, \Delta_j \in \{0, 1\}} \left( E_{\text{ckpt}}((1 - \Delta_j) i_c + \Delta_j j, j) \right.$$

$$+ (\alpha (R + (v_j - v_{i_c}) + \Delta_j C) + \beta R + \gamma) \cdot \sum_{k=i_j+1}^{n} f_k \left. \right),$$_{\text{otherwise}}$

The optimal solution can be computed in $O(n^3)$ time.

Intuitively, $i_c$ denotes the index of the last checkpointed value, while $i_i$ denotes the index of the last value that was tried before we try the next one with index $j$. Here, $\Delta_j$ indicates whether the value $v_j$ will be checkpointed or not. The optimality is proven by induction on the index of the last checkpointed reservation. The proof is again omitted but available in the companion report [9].

C. Approximation algorithm for continuous distributions

In this section, we provide an approximation algorithm of the optimal strategy for continuous distributions with bounded support $[a, b]$, where $a \geq 0$ and $b$ is finite. Because we model job execution times, it is natural to truncate continuous distributions whose support is $[0, \infty]$ such as an Exponential or LogNormal distribution, say, to a bounded support $[a, b]$.

The result for continuous distribution is particularly important: we have shown in recent work [10] that continuous distributions gave strategies that allowed using small data
samples to find an efficient strategy. Here, it returns an arbitrarily good quality solution with low complexity.

More precisely, let $X$ be a continuous random variable defined on $[a, b]$ modeling the probability distribution $\mathcal{D}$, where $0 \leq a < b$, with CDF $F$ and PDF $f$. Theorem 3 shows that Algorithm 1 computes a close-to-optimal strategy for STOCHASTIC. Before stating Theorem 3 we start with a lemma:

**Algorithm 1** DYN-PROG-COUNT($X, \varepsilon$)

1. Let $[a, b]$ be the domain of $X$, with $0 \leq a < b$
2. $c_0 = 3(b-a) \min \left( \frac{1}{\varepsilon}, \frac{1}{\beta} \right)$
3. $n \sim [c_0/\varepsilon]$
4. Define the discrete distribution $Y_n \sim (v_i, f_i)_{i=1\ldots n}$ s.t.
\[
\begin{cases}
  v_i = a + i \cdot \frac{b-a}{n} \\
  f_i = \frac{F(v_i) - F(v_{i-1})}{X \leq v_i} \quad \text{for } 0 \leq i \leq n \\
  f_i = F(v_n) - F(v_{i-1}) \quad \text{for } 1 \leq i \leq n
\end{cases}
\]
5. $S^o_{dp} \leftarrow$ Optimal strategy for $Y_n$ (Theorem 2)
6. return $S^o_{dp}

**Lemma 1.** Given a random variable $X$ and a strategy $\mathcal{S} = \{(t_1, \delta_1), \ldots, (t_i, \delta_i)\}$, if there exists an index $i_0 > 1$ such that $t_1 < \cdots < t_{i_0-1} < \min(R_i, \varepsilon)$, then the strategy $\mathcal{S} = \{\min(R_i, \varepsilon), 0\}, (t_1, \delta_1), \ldots, (t_i, \delta_i)\}$ satisfies:
\[
\mathbb{E}(\mathcal{S}(X)) \leq (1+\varepsilon) \cdot \mathbb{E}(\mathcal{S}(X))
\]

Intuitively, this lemma states that restricting strategies whose first reservation length is at least $\min(R, \varepsilon)$, increases the cost by at most a factor of $1+\varepsilon$.

**Proof.** Consider a strategy $\mathcal{S} = \{(t_1, \delta_1), \ldots, (t_i, \delta_i)\}$ for a random variable $X$, such that there exists an index $i_0 > 1$ with $t_1 < \cdots < t_{i_0-1} < \min(R_i, \varepsilon)$, and define strategy $\tilde{\mathcal{S}} = \{\min(R_i, \varepsilon), 0\}, (t_1, \delta_1), \ldots, (t_i, \delta_i)\}$. From Equation (7), we have $\mathbb{E}(\tilde{\mathcal{S}}(X)) \geq \mathbb{E}(\mathcal{S}(X)) + \alpha(t_1 + C_i)$, and $\mathbb{E}(\tilde{\mathcal{S}}(X)) - \mathbb{E}(\mathcal{S}(X)) = (\alpha(t_1 + C_i) - \min(R_i, \varepsilon)) \cdot \mathbb{P}(X > t_{i_0-1}) + \sum_{i=0}^{i_0-1} (\alpha(t_1 + C_i) - \min(R_i, \varepsilon)) \cdot \mathbb{P}(X > t_{i_0-1}) + \sum_{i=0}^{i_0-1} (\alpha(t_1 + C_i) - \min(R_i, \varepsilon)) \cdot \mathbb{P}(X > t_{i_0-1})$.

We have shown that $C_i = \tilde{C}_i$, $\forall i \geq i_0$. Now we show that $\forall i \geq i_0$, $W_i = \tilde{W}_i$. We consider two cases: (i) the last checkpoint before $t_i$ was done during $t_j$ with $j < i_0$ or there was no checkpoint before $t_i$. In this case, $W_i = \tilde{W}_i$; (ii) the last checkpoint before $t_i$ was done during $t_j$ with $j < i_0$ in $\mathcal{S}$, and there was no checkpoint done in $\tilde{\mathcal{S}}$ before $t_i$. In this case, we have $W_i = R_i + (t_i - t_j) + \delta_i C_i$.

Finally, $\mathbb{E}(\mathcal{S}(X)) \geq (\alpha + \beta) \mathbb{E}[X] + \gamma$, because this is the cost of an omniscient strategy that makes a single reservation of exactly the right size for each job. Therefore, we obtain $\mathbb{E}(\mathcal{S}(X)) - \mathbb{E}(\mathcal{S}(X)) \leq \varepsilon \cdot \mathbb{E}(\mathcal{S}(X)).$

**Theorem 3.** Given a continuous random variable $X$ on the domain $[a, b]$, where $0 \leq a < b$, and given a constant $\varepsilon > 0$, DYN-PROG-COUNT($X, \varepsilon$) is a $1 + \varepsilon$-approximation algorithm for STOCHASTIC and executes in time $O(1/n)$.}

**Proof.** Given a continuous random variable $X$ of support $[a, b]$, we define the discrete random variable $Y_n \sim (v_i, f_i)_{i=1\ldots n}$ as stated in Equation (9) of Algorithm 1. Let $S^o_{dp} = \{(t_i^o, \delta_i^o)\}_{1 \leq i \leq |S^o_{dp}|}$ denote the optimal solution for $X$, and let $S^o_{dp}$ be the optimal solution for $Y_n$ returned by Theorem 2.

We want to show that $\mathbb{E}(S^o_{dp}(X)) \leq (1+\varepsilon) \cdot \mathbb{E}(S^o_{dp}(X))$. In order to do so, we construct two intermediate strategies $S^o_{sp}$ and $S^o_{dp}$ as follows.

First, $S^o_{sp} = \{(t_i^o, \delta_i^o)\}_{i=1\ldots n}$ is constructed in such a way that if $t_i^o \geq \min(R_i, \varepsilon)$, then $S^o_{sp} = S^o_{dp}$ otherwise we construct $S^o_{sp}$ from $S^o_{dp}$ by Lemma 4 below, with the value $\delta_i^o$. Then, according to Equation (5) we have:
\[
\mathbb{E}(S^o_{sp}(X)) \leq (1 + \varepsilon) \cdot \mathbb{E}(S^o_{sp}(X))
\]

Second, $S^o_{dp} = \{(t_i^o, \delta_i^o)\}_{1 \leq i \leq |S^o_{dp}|}$ (hence $|S^o_{dp}| = |S^o_{dp}|$), is such that for $1 \leq i \leq |S^o_{dp}|$, we let $t_i^o, \delta_i^o = (v_i, \pi^o(i))$. Here, we use the sequence $(v_i)_{i=1\ldots n}$ from Equation (9), and the function $\pi^o$ defined by $v_{\pi^o(i)-1} < t_i^o \leq v_{\pi^o(i)}$. In other words, for each reservation, $S^o_{dp}$ chooses the first discrete value larger than or equal to the corresponding one chosen by $S^o_{sp}$, and makes the same checkpointing decision.

**Lemma 2.** $\mathbb{E}(S^o_{dp}(X)) \leq (1 + \varepsilon) \cdot \mathbb{E}(S^o_{sp}(X))$.

**Proof.** We use the notations $T_i^o, R_i^o, C_i^o, W_i^o$ for the parameters of $S^o_{dp}$ and $T_i^o, R_i^o, C_i^o, W_i^o$ for the parameters of $S^o_{dp}$.

From Equation (4), we see that, for $1 \leq i \leq |S^o_{tp}|$, we have $\delta_i^o = \delta_i^o, R_i^o = R_i^o, C_i^o = C_i^o,$ and $W_i^o - W_i^o = T_i^o - T_i^o$. In addition, if $\sigma^o(i)$ (resp. $\sigma^o(i)$) is the index of the last checkpoint before $t_i^o$ (resp.) $t_i^o$, then $\sigma^o(i) = \sigma^o(i)$, and,
\[
|T_i^o - T_i^o| = \left( t_i^o - t_i^o \right) - \left( t_i^o - t_i^o \right) \leq \left( v_{\pi^o(i)} - v_{\pi^o(i)} \right) - \left( t_i^o - t_i^o \right) \leq \left( v_{\pi^o(i)} - t_i^o \right) \leq \left( v_{\pi^o(i)} - t_i^o \right) \leq \left( v_{\pi^o(i)} - t_i^o \right) \leq \varepsilon \cdot |\frac{a}{n}|
\]
From Equation (7) we have: $\mathbb{E}(S^o_{dp}(X)) = \beta \mathbb{E}[X] + \sum_{i=1}^{|S^o_{sp}|} \left( \alpha W_i^o + \beta (R_i^o + (1 - \delta_i^o) T_i^o - C_i^o - T_i^o) + \gamma \right) \cdot \mathbb{P}(X > t_i^o-1)$.
We first observe that $P(X > t_{i-1}) \leq P(X > t_i)$ because $t_i \geq t_{i-1}$. We can derive that $\mathbb{E}(S_{\text{algo}}(X)) - \mathbb{E}(S_{\text{opt}}^{i/3}(X)) \leq \sum_{i=1}^{\lfloor S/X \rfloor} (\alpha |T_i - T_i| + \beta(1 - \delta_i)|T_i| - T_i - T_i-1)) \cdot P(X > t_i) \leq \alpha \frac{b-a}{n} + \sum_{i=1}^{\lfloor S/X \rfloor} (\alpha + \beta(1 - \delta_i)) \frac{b-a}{n} \cdot P(X > t_i)

We also observe that:

$\mathbb{E}(S_{\text{opt}}^{i/3}(X)) \geq \gamma \cdot P(X > t_i) \cdot \mathbb{E}(X)$

Furthermore, for $1 \leq i \leq \lfloor S/X \rfloor$, we have $W_i \geq R_i + T_i \geq \min(R, a)$, where $\hat{a} = \max(a, \min(R, \mathbb{E}(X)/3))$. This is because either $T_i \geq \hat{a}$ according to Lemma [1] (when there was no checkpoint before $t_i$), or $R_i = R$ (when there was a checkpoint before $t_i$). Therefore:

$\mathbb{E}(S_{\text{opt}}^{i/3}(X)) \geq \min(a, \min(R, \epsilon |X|/3)) \cdot \mathbb{E}(X) = \max(a, \min(R, \epsilon |X|/3))$, $R, C$

One can observe that:

$\frac{|S|}{\epsilon} \cdot \gamma \cdot P(X > t_i) \cdot \mathbb{E}(X) \leq \frac{|S|}{\epsilon} \cdot \mathbb{E}(S_{\text{opt}}^{i/3}(X))$

Using $c_0$ (line 2, Algorithm [1]), we obtain:

$\mathbb{E}(S_{\text{algo}}(X)) - \mathbb{E}(S_{\text{opt}}^{i/3}(X)) \leq c_0 \cdot \mathbb{E}(S_{\text{opt}}^{i/3}(X)) = \frac{\epsilon}{3} \cdot \mathbb{E}(S_{\text{opt}}^{i/3}(X))$

**Lemma 3.** $\mathbb{E}(S_{\text{algo}}^{i/3}(X)) \leq \mathbb{E}(S_{\text{algo}}^{i/3}(X))$

**Proof.** Given any reservation strategy $S = \{(i, \delta_i)\}_{1 \leq i \leq |S|}$ such that $\forall i, t_i \in \{v_1, \ldots, v_n\}$, we show that:

$\mathbb{E}(S(Y_n)) - \mathbb{E}(S(X)) = \beta (\mathbb{E}[Y_n] - \mathbb{E}[X])$

Indeed, for the two distributions $Y_n$ and $X$, the only differences in the cost function are: (i) the expectations $\mathbb{E}[Y_n]$ and $\mathbb{E}[X]$; and (ii) the probability values $P(Y_n > t_i)$ and $P(X > t_i), \forall i$. But if $t_i \in \{v_1, \ldots, v_n\}$, we have:

$P(Y_n > t_i) = P(Y_n > v_k) = P(Y_n \in \bigcup_{j=k+1}^{n} \{v_j\}) = \sum_{j=k+1}^{n} P(Y_n = v_j)$

$P(X > v_k) = P(X > t_i)$

We obtain that:

$\mathbb{E}(S(Y_n)) - \mathbb{E}(S(X)) = \beta (\mathbb{E}[Y_n] - \mathbb{E}[X])$

We apply this result to both $S_{\text{algo}}^{i/3}$ and $S_{\text{algo}}^{i/3}$ and derive that:

$\mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) - \mathbb{E}(S_{\text{algo}}^{i/3}(X)) = \mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) - \mathbb{E}(S_{\text{algo}}^{i/3}(X))$

or equivalently,

$\mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) - \mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) = \mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) - \mathbb{E}(S_{\text{algo}}^{i/3}(X))$

$L_{\text{opt}}$ is optimal for $Y_n$, hence $\mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) - \mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) \leq 0$.

And finally, the result $\mathbb{E}(S_{\text{algo}}^{i/3}(X)) - \mathbb{E}(S_{\text{algo}}^{i/3}(X)) \leq 0$.

Combining Lemma [2], Lemma [3] and Equation [10], we get:

$\mathbb{E}(S_{\text{opt}}^{i/3}(X)) \leq \mathbb{E}(S_{\text{algo}}^{i/3}(Y_n)) \leq \left(1 + \frac{\epsilon}{3}\right) \cdot \mathbb{E}(S_{\text{algo}}^{i/3}(X))$

$\leq \left(1 + \frac{\epsilon}{3}\right) \cdot \left(1 + \frac{\epsilon}{3}\right) \cdot \mathbb{E}(S_{\text{opt}}^{i/3}(X))$

which concludes the proof of Theorem [3].

**D. Extensions**

All the results presented in Sections [III-A] to [III-C] namely the cost model (Theorem [1]), the optimal algorithm for discrete distributions (Theorem [2]), and the approximation algorithm for continuous distributions with bounded support (Theorem [3]), can be extended to some variants of the problem where the checkpoint strategy is determined a priori.

Indeed, there are two important and natural variants to consider: strategies where no reservation is checkpointed, and strategies where all reservations are checkpointed. The former variant (called No-CKPT) was studied in our previous work [3], where we derived an optimal algorithm for discrete distributions with reduced time complexity $O(n^2)$ instead of $O(n^3)$ as in Theorem [2]. The latter variant (called ALL-CKPT) also admits an optimal dynamic programming algorithm of reduced time complexity $O(n^2)$ that can be found in the companion report [9].

**IV. PERFORMANCE EVALUATION**

In this section, we evaluate the performance of the different algorithms in simulation. In the following, performance stands for the expected cost of each algorithm under various job execution time distributions, C/R overheads and cost functions. We use jobs that follow a wide range of usual probability distributions as well as a distribution obtained from traces of a real neuroscience application. The code for this section is publicly available on https://gitlab.inria.fr/vhonore/ckpt-for-stochastic-scheduling.

**A. Evaluated algorithms**

In addition to the algorithms presented in Section [III] we propose a periodic heuristic for the case of bounded distributions. This strategy, described in Algorithm [2], is a natural policy, where successive reservations differ in length by a constant amount of time $T$, called the period. A checkpoint is performed at the end of each period. Hence, the value of $W_i$ associated with each $t_i$ is constant in this strategy. The algorithm specifies the number of chunks $\tau$ in the domain $[a, b]$ of the bounded distribution, thus the period can be computed as $T = \frac{b-a}{\tau}$. Note that for this policy, one can derive optimal strategies for some distributions (such as uniform distributions [9]).

Overall, we evaluate five different algorithms from the following two sets of strategies:

- **DYN-PROG-COUNT:** This set includes Algorithm [1] and its ALL-CKPT and NO-CKPT variants described in Section [III-D].
- **ALL-CKPT-PER:** This set includes Algorithm [2] and its NO-CKPT-PER counterpart where checkpointing is not allowed (i.e., $\delta_i = 0, \forall i$).
Algorithm 2 ALL-Ckpt-Per($X, \tau$)
1: Let $[a, b]$ be the domain of $X$, and let $T = \frac{b-a}{\tau}$
2: $(t_i, \delta_i) = \begin{cases} (a + i \cdot T, 1) & \text{for } i = 1, 2, \ldots, \tau - 1 \\ (b, 0) & \text{for } i = \tau \end{cases}$
3: return $\mathbf{q}_{\text{target}} \leftarrow (t_i, \delta_i)_{1 \leq i \leq \tau}$

Table I: Probability distributions and parameter instantiations.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PDF $f(x)$</th>
<th>Instantiation</th>
<th>Support (in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential ($\lambda$)</td>
<td>$\frac{1}{\lambda}e^{-x/\lambda}$</td>
<td>$X \sim 1.0h^{-1}$</td>
<td>$t \in [0, \infty)$</td>
</tr>
<tr>
<td>Weibull ($\lambda, k$)</td>
<td>$\frac{\lambda}{\kappa}e^{-((x/\lambda)^{\kappa})}$</td>
<td>$\lambda = 1.0h, k = 0.5$</td>
<td>$t \in [0, \infty)$</td>
</tr>
<tr>
<td>Gamma ($\alpha, \beta$)</td>
<td>$e^{-\mu} \frac{\mu^{a-1}}{\Gamma(a)}$</td>
<td>$\alpha = 2.0, \beta = 2.0 - 1$</td>
<td>$t \in [0, \infty)$</td>
</tr>
<tr>
<td>LogNormal ($\mu, \kappa$)</td>
<td>$\frac{1}{x \sqrt{2\pi} \sigma} e^{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}}$</td>
<td>$\mu = 1.5h, \kappa = 3.0$</td>
<td>$t \in (0, \infty)$</td>
</tr>
<tr>
<td>Pareto ($\alpha, \beta$)</td>
<td>$\frac{\alpha^{-1} \beta^\alpha}{(x + \beta)\left[1 + \frac{\beta}{x}\right]}$</td>
<td>$\alpha = 1.0h, \beta = 2.0h$</td>
<td>$t \in [\alpha, \infty)$</td>
</tr>
</tbody>
</table>

Distributions with finite support

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Instantiation</th>
<th>Support (in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truncated Normal ($\mu, \kappa, a, b$)</td>
<td>$\frac{1}{\sqrt{2\pi}\beta^{\frac{1}{2}}-\frac{1}{1-a}} e^{-\frac{(x-a)^2}{2\beta^2}}$</td>
<td>$\mu = 2.0h, \kappa = 2.0h, a = 2.0h, b = 20.0h$</td>
</tr>
<tr>
<td>Uniform ($\alpha, \beta$)</td>
<td>$\frac{1}{\beta - \alpha}$</td>
<td>$\alpha = 0.0h, \beta = 20.0h$</td>
</tr>
<tr>
<td>Beta ($\alpha, \beta$)</td>
<td>$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1} (1-x)^{\beta - 1}$</td>
<td>$\alpha = 2.0, \beta = 2.0$</td>
</tr>
<tr>
<td>Bounded Pareto ($L, H, \alpha$)</td>
<td>$\frac{\alpha L^{\alpha - 1}}{L^{\alpha} - H^{\alpha - 1} - (x^{\alpha} - H^{\alpha - 1})}$</td>
<td>$L = 1.0h, H = 20.0h, \alpha = 2.0$</td>
</tr>
</tbody>
</table>

B. Evaluation methodology

We evaluate the performance using two scenarios, both based on the Reserved Instance pricing scheme in AWS [2], where the user pays exactly what is requested. In the evaluation, we set $\alpha = 1, \beta = \gamma = 0$:

- **Scenario 1** (Section IV-C): We consider nine usual probability distributions, five of which have infinite support (Exponential, Weibull, Gamma, LogNormal, Pareto) and four have finite support (Truncated Normal, Uniform, Beta, Bounded Pareto). Table I lists all distributions used in simulation with the instantiations of their parameters for evaluation. The first five distributions are truncated and fed as input to Algorithm 1.

To do so, we set the upper bound of the infinite support to $b = 3600 - v$, where $Q(x) = \inf\{t \mid F(t) \geq x\}$ is the quantile function and $v$ is a small constant. In our simulation, we set $v = 10^{-7}$. During the discretization procedure in Algorithm 1, we then normalize the probabilities of all discrete values so that they sum to 1. We set $C = R = 360$ seconds (0.1 hour). This checkpointing cost is extracted from [18] and corresponds to an average checkpointing duration, where an optimistic one is 60 seconds and a pessimistic one is 600 seconds. We further discuss the impact of the checkpointing cost on the performance.

- **Scenario 2** (Section IV-D): In this scenario, we consider execution traces of a real neuroscience application, and fit a LogNormal distribution to its execution times. To further evaluate the robustness of the algorithms, we perturb the parameters of the fitted distribution by varying its mean and standard deviation, and show the impact on the performance.

Additional simulations with different cost models where $\beta \neq 0$ are available in [9], with similar trends as the results presented below.

C. Results for Scenario 1

We first evaluate the performance of DYMPRO-COUNT compared to the other strategies, as a function of the values of $R$ and $C$. Figure 5a presents the performance of these strategies normalized to that of DYMPRO-COUNT (black line for $y = 1.0$) for the Exponential distribution (Figure 5a) and Boundary Pareto distribution (Figure 5b). The results are similar for other distributions [9]. We use $\varepsilon = 0.1$ for DYMPRO-COUNT and its variants. Regarding periodic strategies, we choose the best value for the number of chunks $\tau$ in $[1, 1000]$. Not surprisingly, we can observe that when $C$ and $R$ are small, the best result is to use the ALL-Ckpt strategy while they are large, one should use the No-Ckpt strategy. There exist thresholds on the sizes of $C$ and $R$ where DYMPRO-COUNT uses a mix of checkpointed and not checkpointed reservations. In that case, the gain of using DYMPRO-COUNT can be up to 10% compared with its variants. An interesting future research direction is to find properties on those thresholds as a function of the probability distribution. Finally, one should observe that the gain achieved with DYMPRO-COUNT compared to the best periodic solution is in general even larger than 10%. The exception is for the Exponential distribution, where one can show [9] that ALL-Ckpt and its periodic counterpart are identical. This is due to the memoryless property of the Exponential distribution.

We then study the impact of $\varepsilon$ on the performance of DYMPRO-COUNT (DPC) when $R = C = 360$s. The idea is that when $\varepsilon = 1$, this theoretically guarantees that the performance is at most twice ($= 1 + \varepsilon$) that of the optimal, but in practice it can be a lot better. In Figure 6, we study the performance of DYMPRO-COUNT for various values of $\varepsilon$ for distributions of Table I. All performance are normalized by DYMPRO-COUNT for $\varepsilon = 0.1$. We see that in practice, the convergence to the lower bound in performance is fast. Indeed, for $\varepsilon = 1$, almost all distributions already reach convergence, except for Weibull and Pareto (which have a much larger domain of definition and specific properties). For those

\[ \frac{1 - F(x + y)}{1 - F(x)} \rightarrow 1 \text{ when } x \rightarrow \infty, \quad \forall y > 0. \]
distributions, we see that they converged for $\varepsilon = 0.1$. For this experiment, the number of chunks $n$ in DYN-PROG-COUNT varies between 50 to 1000 depending on the distribution and value of $\varepsilon$, showing the practicality of DYN-PROG-COUNT for the target distributions.

Our final evaluation for this scenario is a study of the impact of the period size. So far, we have always chosen the period minimizing the objective functions. Table I shows the performance of both variants of the periodic algorithms, ALL-Ckpt-PER and NO-Ckpt-PER, normalized by that of DYN-PROG-COUNT ($\varepsilon = 0.1$), when $C = R = 360s$. For each distribution, the second column shows the best period found when $\tau$ varies from 1 to 1000 (with its associated cost normalized by that of DYN-PROG-COUNT), and the other columns present results for specific values of $\tau$ in that interval. As observed before, ALL-Ckpt-PER is in general not able to match DYN-PROG-COUNT (except for some distributions). In addition, we see that NO-Ckpt-PER performs even worse than ALL-Ckpt-PER. The reason is that the checkpointing cost is relatively low in this setup, so it is preferable to checkpoint more often than never. Finally another observation is that a wrong period size can significantly deteriorate the performance of the periodic algorithms.

D. Results for Scenario 2

We now present the simulation results for a probability distribution fitted to the execution times from the traces of a real neuroscience application (a code for structural identification of orbital anatomy) extracted from the Vanderbilt’s medical imaging database [15]. Figure 4 shows the execution traces of the application and its fitted LogNormal distribution. Figure 4 presents the performance of different algorithms for this fitted distribution. To evaluate the robustness of algorithms, we also vary the original mean $\mu_o$ (Figure 7a) or standard deviation $\sigma_o$ (Figure 7b) of the distribution from their original values. For readability, all axis are in logscale.

We fix the checkpointing cost to $C = R = 600$ seconds and $\varepsilon = 1.0$. For periodic strategies, we use similar brute-force procedure as Scenario 1 to find the period that performs best.

The expected costs of the algorithms are normalized by that of an omniscient scheduler (blue dashed line), which knows the execution time $t$ of a job a priori, and thus would pay the minimum possible cost by making a single reservation of length $t_1 = t$. We observe that DYN-PROG-COUNT always gives the best performance. As previously seen, the checkpointing cost influences the performance of NO-Ckpt and ALL-Ckpt with regard to DYN-PROG-COUNT. In this setup, since $C = R = 600$ seconds is a value low enough to allow for checkpointing all reservations, the performance of DYN-PROG-COUNT and ALL-Ckpt are the same and outperforms NO-Ckpt by a wide margin. Simulation with other C/R values can be found in [9] and show similar trends. As for the periodic algorithms, ALL-Ckpt-PER has better performance than NO-Ckpt-PER. However, both algorithms have worse performance than DYN-PROG-COUNT. The results demonstrate the robustness of DYN-PROG-COUNT for a practical application with different distribution parameters.

Finally, when the ratio $\mu/\sigma$ is large (either by increasing the mean ($\mu/\mu_o$) large), or decreasing the standard deviation ($\sigma/\sigma_o$ small)), the solutions converge to the omniscient scheduler. This could be expected, since in this case the variability becomes negligible and the job behaves similarly to a deterministic job.

V. Experiments

In this section, we conduct real experiments on an HPC platform by using three stochastic neuroscience applications. The focus is to study the performance of different reservation and checkpointing strategies when scheduling multiple jobs in a shared HPC execution environment.

A. Experimental setup

The chosen neuroscience applications are described in Table III along with their execution characteristics, which are extracted from the Vanderbilt’s medical imaging database [15]. In particular, the walltime distributions are obtained by fitting traces of execution times, while the checkpointing/restart costs are obtained by analyzing and averaging memory footprints. Note that, for these applications, restart costs ($R$) differ from checkpointing costs ($C$) and depend upon the time-steps at which they are taken. We focus on the evaluation of the
following two different sets of strategies:

- An HPC-for-neuroscience strategy (called HPC in Section [V-B]), which uses the average of the last 5 runs as the initial reservation length and then increases it by 50% for each subsequent reservation. This strategy is currently used by the MASi group [22] at Vanderbilt to handle stochastic neuroscience applications.

- Our proposed DYNS-PG-COUNT strategy and its ALL-CKPT variant.

We ran the experiments on a 256-thread Intel Processor (Xeon Phi 7230, 1.30GHz) while submitting jobs through the Slurm scheduler. All three neuroscience applications are sequential (i.e., use a single hardware thread) and perform some medical imaging analysis. The variation in execution time is due to the different characteristics of the input data. However, as we do not have access to the raw input images, we used the information in the logs to simulate the characteristics of the input data, thereby forcing a job to run for a certain walltime and saving a specific amount of data for the checkpoints. The platform under study obeys the RESERVATIONONLY cost model, with $\alpha = 1$, $\beta = \gamma = 0$. In each experiment, we submitted 500 total jobs, and recorded the completion time of each of them. We use the average job stretch (defined as the ratio between the total execution time of a job and its actual walltime) to show the individual job performance, and use the utilization (defined as the ratio between the sum of all job walltimes and the total time required to execute them) to show the performance of the system for the whole job set. By experimenting on a real system, we investigate the robustness of our strategy: 1) when multiple applications of the same type are running concurrently (and read/write times vary due to congestion while accessing I/O and/or due to application interference); 2) when the C/R costs vary depending upon which time-steps get a checkpoint/restart (i.e., different values for different reservations); and 3) when running multiple job types concurrently.

### B. Experimental results

Figure 8 shows the performance of the three strategies when submitting 500 jobs from each application to the Slurm scheduler. In this experiment, we manually force the C/R costs to be the same (as in Table III) for each strategy, in order to study the impact of application interference and runtime system’s performance variability on our model. The findings are consistent with the simulation results (in Section [V]), showing that DYNS-PG-COUNT performs better than its ALL-CKPT variant in terms of both system utilization and average job stretch using all three applications. Moreover, both algorithms outperform the simple HPC strategy.

Depending on when the checkpoint is being taken, the checkpoint size and thus the time to save and restore the application can vary. Figure 9 shows the results when the C/R costs could vary for different reservations. Based on the log traces of these three applications, we noticed that their memory footprints can vary by as much as 30% depending on when the checkpoint is taken (e.g., the checkpoint time can vary between 80 and 110 seconds for Qball). Our experiment generates random checkpoint sizes using a uniform distribution with the mean given by the average checkpoint size from the traces, and forces the application to read/write the corresponding amount at the beginning/end of the execution. In this experiment, we assume that the checkpointing time is included in the request time and is never responsible for applications exceeding their allocated time. While the DYNS-PG-COUNT solution is computed using the average C/R costs presented in Table III the experimental results show that its performance is robust up to 15-20% variability in the C/R costs. Moreover, the average job stretch appears to be even more stable than the utilization, suggesting that most of the submitted jobs are not impacted by the fluctuation in the C/R costs.

If application-level checkpointing is used, the application is usually aware of the checkpoint size, thus the checkpointing process can start before the reservation is over. The subsequent submissions can easily adapt to this deviation with the first checkpoints that are smaller than the one used to compute the sequence (this is the case for Figure 9). For system-level checkpointing, the application footprint usually remains similar throughout the execution of the application. In case

---

**Table III: Characteristics of the chosen neuroscience applications.**

<table>
<thead>
<tr>
<th>Application type</th>
<th>Wannier distribution</th>
<th>t</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diffusion model fitting (Qball)</td>
<td>Gamma ($\alpha = 1.15$, $\beta = 4.3$, $\gamma = 1465$, 407s)</td>
<td>80s</td>
<td>40s</td>
</tr>
<tr>
<td>Diffusion model fitting (SD)</td>
<td>Gamma ($\alpha = 1.15$, $\beta = 4.3$, $\gamma = 1465$, 407s)</td>
<td>25min</td>
<td>10min</td>
</tr>
<tr>
<td>Functional connectivity analysis (PCA)</td>
<td>Gamma ($\alpha = 1.15$, $\beta = 4.3$, $\gamma = 1465$, 407s)</td>
<td>150s</td>
<td>100s</td>
</tr>
</tbody>
</table>

---

**Table II: Expected cost of ALL-CKPT-PER and NO-CKPT-PER, normalized by DYNS-PG-COUNT(X, 0.1). C = R = 360s.**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Best $\gamma$</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 200$</th>
<th>$\gamma = 400$</th>
<th>$\gamma = 600$</th>
<th>$\gamma = 800$</th>
<th>$\gamma = 1000$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
<td>23 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
<td>91 (1.00)</td>
</tr>
</tbody>
</table>

---

**Fig. 8: Utilization (higher is better) and average job stretch (lower is better) for DYNS-PG-COUNT, ALL-CKPT and the HPC strategies.**
the checkpointing time is causing the application to exceed the reserved time, the submission will fail and subsequent submissions can take this into account by adding the wasted time. The limitation of our method is visible for applications with large variability in checkpointing size, which can be due to multiple factors, either within the application that presents different memory footprints throughout its execution, or by system-level causes, such as I/O congestion or failures. Such large variability in checkpointing size compared to what is used to compute the reservation sequence can result in worse performance when using our method, and the classic HPC model would be preferred in this case. We are currently investigating methods to incorporate variation of checkpointing size into the computation of the optimal reservation sequence, by either using historic information or adapting the subsequent request times based on the sizes of previous checkpoints. We plan to further analyze variable C/R times in the future.

Finally, we conduct experiments in a more realistic scenario by running all three applications simultaneously, and investigating the impact on the different strategies. Specifically, we submitted a total of 500 jobs (100 from Qball, and 200 each from SD and FCA), and kept the C/R costs constant across different reservations to study the sole impact of having several application types executing concurrently. We recorded the utilization and average job stretch when using DYN-PROG-COUNT compared to the HPC strategy for 10 different runs choosing different instances from the traces each time. The results are presented in Table IV. We see that DYN-PROG-COUNT improves both utilization and average job stretch by 10% on average, and by up to 20% depending on the instances submitted. Overall, these results again illustrate the robustness of our algorithm and confirm its benefit for scheduling stochastic applications on reservation-based platforms, as long as checkpoint costs remain constant for each application.

VI. RELATED WORK

We review some related work on reservation-based scheduling and checkpointing in HPC and cloud systems, as well as some prior work on dealing with stochastic applications.

a) Reservation-based scheduling: Batch schedulers are widely adopted by many resource managers in HPC systems, such as Slurm, Torque and Moab. Most batch schedulers use resource reservation in combination with backfilling [23], [25], [27], and rely on users to provide accurate estimates for the walltimes of the submitted jobs. While this works for applications with deterministic resource needs, it can cause resource over-estimation or under-estimation for stochastic jobs with large variations in the walltime, thus degrading system and/or application performance [12], [29].

Clusters of commodity servers that use big-data frameworks such as MapReduce [7] and Dryad [19] offer alternative solutions to running HPC workloads. Schedulers for these frameworks such as YARN [28] and Mesos [16] offer distinct features (e.g., fairness, resource negotiation) to manage the workloads, but they generally also require accurate information regarding the applications’ resource demands.

Cloud computing platforms such as Amazon AWS [2] and Google GCP [14] have emerged as another option for executing HPC applications, with a variety of pricing and reservation schemes. Both on-demand and reservation models are available with the latter typically offering a lower price. Several works [1], [5], [8], [31] have studied the pricing strategies for platform providers, as well as delay modeling and cost evaluation for the users.

b) Stochastic scheduling and checkpointing: Many prior works have considered stochastic scheduling for jobs with execution time uncertainty. Most research in this paradigm (e.g., [4], [13], [21], [24], [26]) assumes that the execution time of a job follows a known probability distribution and aims to optimize the expected response time or makespan for a set of jobs under various distributions. Most of them, however, do not consider the problem in the context of reservation-based scheduling. In our prior work [3], we have proposed near-optimal reservation strategies for a single job in both HPC and cloud systems. The work was later extended to scheduling a set of stochastic jobs, both sequential and parallel, using backfilling in a reservation-based environment [11], [12].

Another approach to coping with stochastic applications and/or platform unavailability is through checkpoint-restart [17], [30]. To ensure the robustness of the execution, the application’s state is periodically checkpointed and in case of interrupt (due to either insufficient reservation or platform failure), the application can be recovered from the last saved checkpoint. In the context of fault tolerance, a lot of work (e.g., [6], [17], [32]) has been devoted to deriving the optimal checkpointing interval that minimizes the checkpointing size.

<table>
<thead>
<tr>
<th>DYN-PROG-COUNT</th>
<th>HPC</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization</td>
<td>Avg Stretch</td>
<td>Utilization</td>
</tr>
<tr>
<td>67</td>
<td>2.04</td>
<td>55</td>
</tr>
<tr>
<td>75</td>
<td>1.74</td>
<td>62</td>
</tr>
<tr>
<td>62</td>
<td>2.08</td>
<td>55</td>
</tr>
<tr>
<td>71</td>
<td>1.88</td>
<td>64</td>
</tr>
<tr>
<td>83</td>
<td>2.19</td>
<td>56</td>
</tr>
<tr>
<td>71</td>
<td>1.74</td>
<td>64</td>
</tr>
<tr>
<td>75</td>
<td>1.51</td>
<td>68</td>
</tr>
<tr>
<td>68</td>
<td>2.09</td>
<td>65</td>
</tr>
<tr>
<td>61</td>
<td>2.24</td>
<td>60</td>
</tr>
<tr>
<td>77</td>
<td>1.96</td>
<td>75</td>
</tr>
</tbody>
</table>

Fig. 9: Utilization and average job stretch for each application when varying the C/R costs by different percentages (0 to 30%) using the DYN-PROG-COUNT and HPC strategies.
checkpointing costs exhibit a large variability. This would alleviate the limitation of our approach when strategies that will be more robust than our current solutions is to incorporate non-constant checkpointing costs into the for a given application profile. Another interesting direction is to always (or never) checkpoint the reserva-


Altogether, we have demonstrated the effectiveness of these new solutions in comparison with classic strategies. Hopefully, these results should help to convince HPC users and system administrators that significant improvements, in terms of both system and application performance, can be achieved by using a well-chosen reservation sequence rather than a unique reservation of maximum length (the current standard policy).

For future work, we are interested in analytically quantifying the critical checkpointing cost, below (or above) which the best strategy is to always (or never) checkpoint the reserva-

resources were provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l’Adour.

**REFERENCES**


[22] B. Landman. Medical-image Analysis and Statistical Interpretation (MASI) Lab. [https://my.vanderbilt.edu/masi/](https://my.vanderbilt.edu/masi/)
[29] O. Weidner, M. Atkinson, A. Barker, and R. Filgueira Vicente. Re-

overhead or resource waste.

In this paper, we present strategies that combine reservation and checkpointing for stochastic jobs with known execution time distributions. To the best of our knowledge, this is the first result to provide performance guarantee on the expected execution time while leveraging checkpointing in reservation-based scheduling environment.

**VII. CONCLUSION AND FUTURE WORK**

We have studied the problem of scheduling stochastic jobs running on a reservation-based platform. We presented a model and optimization framework which combine a sequence of reservations with associated checkpointing decisions. We provided an optimal solution via a dynamic programming algorithm in the case of discrete distributions. We also provided approximation scheme for bounded continuous distributions that are arbitrarily close to the optimal. We used both standard distributions and traces from real neuroscience applications to conduct an extensive set of simulations and actual experiments. Altogether, we have demonstrated the effectiveness of these new solutions in comparison with classic strategies. Hopefully, these results should help to convince HPC users and system administrators that significant improvements, in terms of both system and application performance, can be achieved by using a well-chosen reservation sequence rather than a unique reservation of maximum length (the current standard policy).

For future work, we are interested in analytically quantifying the critical checkpointing cost, below (or above) which the best strategy is to always (or never) checkpoint the reservations. This will help to fully characterize the optimal solution for a given application profile. Another interesting direction is to incorporate non-constant checkpointing costs into the optimization problem, in order to design new reservation strategies that will be more robust than our current solutions. This would alleviate the limitation of our approach when checkpointing costs exhibit a large variability.

**Acknowledgments:** We thank the anonymous reviewers for their comments and suggestions. This research was supported in part by the Vanderbilt Institutional Fund. Some of the simulations presented in this paper were carried out using the PlaFIRM experimental testbed, supported by Inria, CNRS (LABRI and IMB), Université de Bordeaux, Bordeaux INP and Conseil Régional d’Aquitaine (see [https://www.plafirm.fr/](https://www.plafirm.fr/)). The remaining simulation resources were provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l’Adour.