資源予約に基づく並列ジョブスケジューリング手法の評価

藤原 一毅　　合田 憲人
東京工業大学

大規模並列計算機上でユーザの指定した時刻にジョブを実行する「資源予約」を実現する場合，予約ジョブと一般のジョブとの混在による影響が問題となる．本稿では，4 種類の予約ポリシーにおける両ジョブの性能を比較，検討する．実験の結果，予約ジョブ優先でスケジューリングを行い，一般ジョブの中断・再開を可能にし，かつ予約ジョブの占有率を低く抑えることにより，予約ジョブと一般ジョブの双方を満足させうることが明らかになった．

Evaluation of Parallel Job Scheduling Strategies with Resource Reservation

Ikki Fujiwara　　Kento Aida
Tokyo Institute of Technology

A “Resource reservation” on a parallel computer system allows users to request when their jobs are going to start. However, performance of job scheduling strategies where these reserved jobs and other ordinary jobs are run on the same systems is not clear. This paper examined the performance of four different reservation policies. The results showed that an administrator should give reserved jobs the highest priority, enable ordinary jobs to be suspended, and keep the number of reserved jobs small in order to get satisfactory performance.

1　はじめに

大規模並列計算機上のジョブスケジューリング技術は，これまでローカルユーザのジョブの実行性能向上を主眼に開発されてきた．しかしながら，今後のグリッドコンピューティングの普及に伴い，複数の並列計算機間で同時刻に同一ユーザのジョブを実行する co-scheduling の必要性が高まり，あらかじめ指定した時刻にジョブを実行する「資源予約」が注目されている．資源予約が可能な並列計算機上では一般のジョブと予約ジョブが混在することになるが，これらのジョブが相互に及ぼしあう影響については，Smith らによって問題提起されているものの [1]，まだ十分に明らかにされていない．本稿では，4 種類の予約ポリシーにおける予約ジョブ・一般ジョブそれぞれの性能をシミュレーションにより評価し，ジョブの混在による性能への影響を明らかにする．

2　システムモデル

本評価で用いたモデルは，性能が均一な m 個のノード (PE) からなる並列計算機上で，空間分割方式により，各ジョブにユーザの指定した PE 数×時間だけ PE を割り当てるものである [2]．ユーザはジョブ投入時に実行開始希望時刻を指定する（予約する）ことが可能であり，予約されたジョブは可能なら希望時刻に，そうでなければ希望時刻以降で可能な最も早い時刻に実行が開始される．そのとき，一般ジョブの保護と中断の有無により，次の 4 種類の予約ポリシーを定義する．ここで「保護」とは，ジョブの割り当て予定時刻を遅延させないことをいう．また「中断」とは，実行中の一般ジョブをサスペンドし，空いた PE に予約ジョブを割り当てることをいう．

- **先着優先（中断なし／中断あり）** 予約ジョブは投入時点ですでに到着していた一般ジョブをすべて保護する．
- **予約優先（中断なし／中断あり）** 予約ジョブは一般ジョブをまったく保護しない．

中断なしの場合，後に予約ジョブが控えているとき，それを遅延させるような一般ジョブの割り当ては行われない．中断ありの場合，1 回の中断における途中結果の保存および再スタートにかかるオーバーヘッドは，文献 [3] の例を参考に，4 秒とした．予約ジョブは中断されない．一般ジョブ同士の割り当て順序は，FCFS，Aggressive Backfilling，ならびに Conservative Backfilling によって定めた．

3　性能評価

本評価では現実のワークロードを模した Feitelson 96 モデル [4] を用い，全ジョブからランダムに r [%] のジョブを抜き出して予約ジョブとした．本稿では r を予約率と呼ぶ．予約ジョブは到着時刻の Mrg [hour] 後を希望時刻とし， Mrg は管理者の定める予約受付時間内で一様分布するものとした．その他の

表 1　実験条件

PE 数	$m = 128$
ジョブ数	$n = 10000$
予約率	$r = 0, 25, 50, 75, 100$ [%]
平均実行時間	1670 [sec]
予約受付時間	$0 \leq Mrg \leq 2$ [hour] ,
	$0 \leq Mrg \leq 24$ [hour]

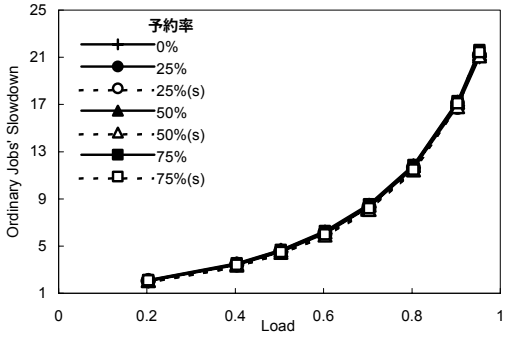

図 1　一般ジョブのスローダウン
（先着優先・中断なし／中断あり (s)）

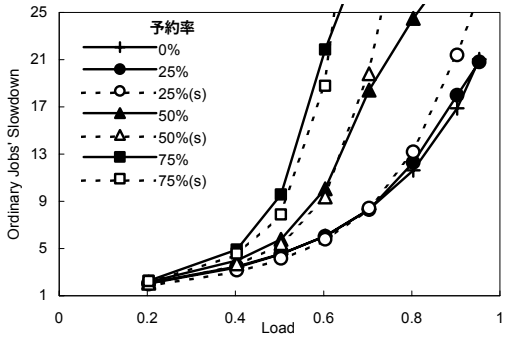

図 2　一般ジョブのスローダウン
（予約優先・中断なし／中断あり (s)）

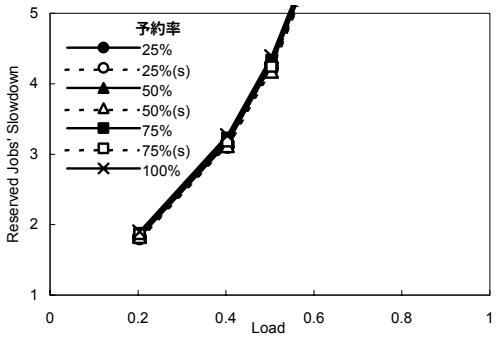

図 3　予約ジョブのスローダウン
（先着優先・中断なし／中断あり (s)）

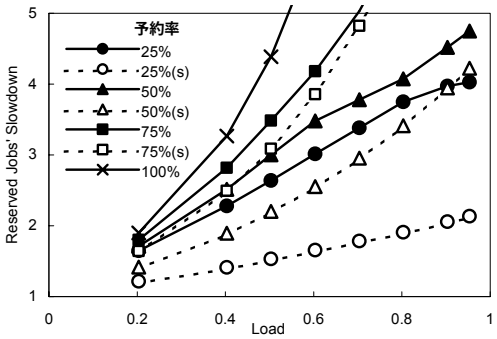

図 4　予約ジョブのスローダウン
（予約優先・中断なし／中断あり (s)）

諸条件は表 1 の通りである．評価指標には次の値を用いる．

$$\text{一般ジョブのスローダウン} = \frac{Fin - Arr}{Run}$$

$$\text{予約ジョブのスローダウン} = \frac{Fin - Req}{Run}$$

ここで *Arr*：到着時刻，*Req*：希望時刻，*Fin*：終了時刻，*Run*：ジョブの正味実行時間 である．なお，ジョブの割り当てにかかるオーバーヘッドは実行時間に対して十分小さいことが別途評価によりわかっているため，本評価では考慮しないこととする．

以上の各手法と予約受付時間の組合せに対して行った実験結果のうち，ここでは Conservative Backfilling，$0 \le Mrg \le 2$ [hour] の結果を示す．

先着優先（図 1・図 3）の場合，一般ジョブは予約による影響をほとんど受けない．一方，予約ジョブはシステム全体の混雑に支配され，予約率を低くしてもスローダウンは改善されない．

予約優先（図 2・図 4）の場合，一般・予約ジョブとも予約率を低くすればそれだけスローダウンが小さくなる．さらに中断ありの場合，予約ジョブは他の一般ジョブからの影響を一切受けず，スローダウンが最小限に抑えられる（図 4）．そのとき一般ジョブには中断・再開のオーバーヘッドが加わるにもかかわらず，スローダウンはさほど増大しない（図 2）．

4　まとめ

本稿では，4 種類の予約ポリシーを用いて予約ジョブ・一般ジョブそれぞれの性能評価を行った．その結果，システムの管理者は (1) 予約優先でスケジューリングを行い，(2) 一般ジョブの中断を可能にし，かつ (3) 予約率を低く保つ ことにより，一般ジョブへの影響を抑えつつ，予約ジョブに最大限のサービスを提供できることが確かめられた．

謝辞　本研究の一部は，日本学術振興会科学研究費補助金（奨励研究 (A)） 課題番号 13780206 による．

参考文献

[1] W. Smith, I. Foster and V. Taylor: “Scheduling with Advanced Reservations”, 14th IPDPS, pp.127-132 (2000).

[2] Kento Aida: “Effect of Job Size Characteristics on Job Scheduling Performance”, Job Scheduling Strategies for Parallel Processing, LNCS 1911, pp.1-17, Springer (2000).

[3] 西岡, 堀, 手塚, 石川: クラスタにおけるコンシステントチェックポイントの実現, JSPP ’98, pp.229-236, 情報処理学会 (1999).

[4] D. G. Feitelson: “Packing schemes for gang scheduling”, Job Scheduling Strategies for Parallel Processing, LNCS 1162, pp.89-110, Springer (1996).