社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－ARC－162 (9)
2005－HPC－101 (9)
2005／3／7

資源予約可能な並列計算機上のジョブスケジューリングに関する研究

田中 徹也[†]　　合田 憲人[†]

現在の並列計算機システムでは，効率性を高めるために資源管理が必要である．この資源管理の 1 つである資源予約という機能がある．これは，ユーザが指定する時間にジョブの実行のために資源の確保を予約するものである．資源予約機能は数種のスケジューリングシステムで提供されているが，この機能下でのジョブスケジューリング手法についての性能評価は未だ詳細に行われていない．そこで本研究では，予約のあるジョブを効率的に実行させると同時に，予約のあるジョブと予約がないジョブとの間の公平性の維持を目標とした資源予約の方法について提案し，シミュレーション実験を通して性能について議論する．

Research on Job Scheduling for a Parallel Computer Supporting Advanced Reservation

Tetsuya TANAKA[†], Kento AIDA[†]

Resource management is necessary to enhance the efficient operation of the current parallel computer system. The advanced reservation, where a user reserves computing resources for a certain time interval to run the user's job, is one of techniques for efficient resource management. While several job scheduling systems provide the advanced reservation mechanism, the detailed performance of the job scheduling scheme with the advanced reservation has not been studied well. This paper proposes the advanced reservation scheme, which enables efficient execution of reserved jobs and preserves the fairness between reserved jobs and non-reserved jobs, and discusses the performance by simulation. The simulation results, where the new metric to evaluate the fairness between reserved jobs and non-reserved jobs is proposed, shows effectiveness of the proposed scheme.

1. はじめに

並列計算機上で特定の時間に計算ノードを占有して利用する要求が増えている. 例を挙げると, グリッドコンピューティングによる位置的に離散した複数の計算ノードの同時利用やイベント利用などである．これに対し，ユーザが指定する時間にジョブ (プログラムに相当) の実行に必要な資源を確保する“資源予約”という機能が考案された．しかし，この機能下のスケジューリング手法について未だ十分な研究が行われておらず, どのようなスケジューリング方針をとればよいかが問題となっている. この問題に取り組んだ研究 [1, 2] では，予約のあるジョブを絶対的に優先することを前提としているが, 予約のあるジョブを予約がないジョブに対してどの程度優先するべきかについては明らかにされていない. また, 予約のあるジョブと予約がないジョブとの間の公平性に関する議論も十分に行われていない.

そこで本稿では, 予約のあるジョブを絶対に優先することなく, 予約がないジョブとで格差の少ない優先付けを行う方法を提案するとともに, 公平性の評価のために, 予約のあるジョブが, 予約がないジョブと比較してどの程度優先されたかを示す指標を提案する．提案手法では，予約のあるジョブの優先度を予約のあるジョブに与える初期優先度と, 予約要求を行った時刻と予約する希望実行開始時刻の差により決定する. 提案する指標は, 予約によって割り当てが遅れた予約がないジョブの量を示すことで, 予約の優先を定量的に評価する.

提案手法を評価するために, シミュレーション実験を行った結果, 提案手法は予約がないジョブに対する影響を低く抑えながら, 予約を絶対優先

† 東京工業大学
Tokyo Institute of Technology

する場合と同程度の割合で予約時間どおりに予約のあるジョブを実行開始することが確認できた.

2. システムモデル

本節ではシミュレーション実験に用いる資源予約付き並列ジョブスケジューラモデルに関して述べる.

2.1. 並列計算機システムのモデル

本稿が対象とする並列計算機システムは複数の計算ノードから構成され, ユーザは 1 以上の計算ノード数を実行に必要とするジョブを投入する. 計算ノードは, すべて同一の性能を有し, 互いに等価接続されていると考える. 各計算ノード上では 1 個のジョブのみが実行され, 複数のジョブが実行されることはない. 割り当て時のノード構成は実行中に変化することはないとする.

2.2. ジョブスケジューラモデル

ジョブスケジューラモデルは空間分割方式 [3]に分類されるモデルを用いる. ユーザが投入するジョブは, 予め実行に必要なノード数と実行時間を指定されている. 指定したノード数と実行時間は実行終了までに変化しない. まず, 投入されたジョブはキューに挿入される. スケジューラは, 指定されたスケジューリング手法に従い, キュー中で実行開始を待つジョブに対し, 必要な計算ノードを実行時間だけ割り当てる. ジョブの実行は中断される場合とされない場合の 2 通りを想定する. ジョブに必要な空き計算ノードが確保できない場合には, 確保できるまでそのジョブをキューで待機させる.

2.3. 資源予約

資源予約では, ユーザが指定した期間に資源を予約して利用する. 予約を行ったジョブは, その予約時間において必要なノード数が確保できるならばその時間に実行を開始し, 確保できない場合は確保できるようになった時点から実行を開始する. 予約の中止や変更は不可能とする.

2.4. ジョブの中断・再開

ジョブの中断は, 実行中のジョブを強制的に中断させる処理である. そのとき, 中断されたジョブはキューに再び戻され, 実行の機会を待つ. 中断時には, ジョブの状態を保存する. 中断されたジョブの再開時には, 中断時に保存された状態へ復帰して実行を行う. 中断と再開それぞれの処理には必要なオーバーヘッドが発生すると考える.

2.5. ワークロード

実験に用いるワークロードは, Feitelson 1996 年モデル (Feitelson 1996 Model) [4,5] をもとに作成した. このモデルではジョブの到着頻度と予約時間は規定されていないため到着頻度と予約時間を与える必要がある. ジョブの到着間隔を与えるために, [1] の方法を用いる. ジョブの到着間隔によってシステムの負荷が定まるものとし, Poisson 到着に従って各ジョブに到着時刻を付与した.

予約時間は, 総ジョブ数の中から決められた割合で予約ジョブをランダムに選び, 予約時間, 即ちジョブの実行を開始する時刻を与えた. 予約時間の指定には, 予約するジョブがキューに挿入されてから予約する実行開始時刻までの期間を決め, その期間からランダムに決定する.

3. 提案するスケジューリング手法

本節では提案するスケジューリング手法の枠組みについて述べるとともに, それを評価する指標について述べる.

3.1. 目標

従来の研究 [1, 2] では, 予約のあるジョブに絶対的な優先度を与えることを前提としている. しかし, 予約のあるジョブに絶対的な優先を与えることはジョブ間の公平性に問題がある. これに対して本稿では, すべてのジョブ間で適当な割り当ての優先の与えることで, 少ない予約の損失で大きな予約の利益を得られるような方法を考える.

3.2. 提案手法の枠組み

本稿が提案するスケジューリング手法は, ジョブ間の割り当て優先度を決定するルール, その優先度から次に割り当てるジョブを選択するルールの 2 つによって定義される.

(a) 優先度決定ルール

提案手法による優先度は, ジョブ間の割り当て優先順位を決定するものである. この優先度は時

予約がないジョブの優先度
＝ 現在時間－投入時間　　(式 1.1)

予約のあるジョブの優先度
＝ジョブの初期優先度 ＋ (現在時間－予約時間)
＋ w ×(予約時間－投入時間)　　(式 1.2)
※ w は係数

図 1．優先度決定ルールの算出式

間ごとに変化し, 実行終了時まで毎単位時間ごとに更新され続ける.

提案手法におけるジョブの優先度は, それぞれ図 1 に示す式 1.1，式 1.2 に従って得られる．図 1 中の投入時間とは, ジョブを投入した時刻である．初期優先度は，予約のあるジョブを優先するために予め与える固定値であり，w は予約した実行開始時刻から投入時間までの時間を重み付けするための係数である.

予約がないジョブの優先度は割り当てまでの待ち時間により決定される. 予約のあるジョブの優先度は, 現在時間が予約時間となったときに初めて発生し, 重み付けされた投入時間と予約時間との時間差，初期優先度と，予約時間からの遅れ 3 つによって得られる.

本手法では，初期優先度及び w の決め方によって, 予約のあるジョブを予約がないジョブに対してどの程度優先すればよいかということ, 希望する実行開始時刻のどの程度前に予約を行えばよいかというという問題について議論できる.

(b)　割り当てルール

これは, 優先度から割り当てるジョブを選択するためのルールである．これには，優先度の順番そのままに割り当てる方法と, 積極的 backfill [3] の 2 種類を考える． backfill とは，優先度の高いジョブを実行するために必要な計算ノードが確保できない場合, これらの優先度の高いジョブの実行開始を遅らせない範囲で, 優先度がより低いジョブを先に割り当てる手法である．積極的 backfill では, backfill を行う基準を, 優先度が，最も高いジョブを遅らせないというものである.

3.3. 評価指標

本稿では，性能評価の指標として， (a) 平均応答時間，(b) 定時予約開始率，(c) 後投入予約先割り当て率 を用いる.

(a) 平均応答時間

平均応答時間は, ジョブの実行開始要求からそのジョブの処理終了までの時間の平均である. 実行開始要求は, 予約がないジョブの場合は投入時間であり, 予約のあるジョブの場合は予約時間である.

(b) 定時予約開始率

定時予約開始率は, 予約のあるジョブが予約時間通りに実行開始した割合である.

(c) 後投入予約先割り当て率

後投入予約先割り当て率は, 本研究で新しく提案する評価指標である.

これは, ある予約のあるジョブが投入される前に到着している予約がないジョブで, 予約時間前に割り当てられていないもののうち, その予約のあるジョブの方が先に割り当てられしまった割合である.

従って本指標では, 予約のあるジョブが挿入されたことにより, 割り当ての機会が遅れた予約がないジョブを調べることで, 予約のあるジョブが優先されることにより実際に得た利益を表す.

4. シミュレーションによる評価実験

本節では実験方法と結果の評価について述べる.

4.1. 実験の条件

以下では，実験を行った条件について，(a) ワークロード，(b) 2 つのルール，(c) 中断・再開，(d) 統計方法 について示す.

(a) ワークロード

対象とする並列計算機の計算ノードの総数は 256，投入する総ジョブ数は 10,000 とした．その他の条件は，表 1 の通りである．表 1 中の予約期間とは, 予約するジョブがキューに挿入されてから予約する実行開始時刻までの期間である. すべての値の組み合わせにつき, 乱数の種を変えて 50 通りのワークロードを生成した.

(b) 優先度決定ルールと割り当てルール

予約に与える優先として試行した初期優先度, w の値の組み合わせは表 2 のとおりである．さらに，予約のあるジョブを絶対優先する場合

表 1．計算ノード数，ジョブの設定条件

ジョブの到着間隔	30, 50, 70, 80, 90 [%]
予約のあるジョブの割合 (ジョブ総数のうち)	10，20 [%]
予約の期間 (ジョブの投入後から)	0-3, 0-12, 0-24, 0-48, 12-24，12-48，24-48 [時間後]

表 2．初期優先度，w の組み合わせ

初期優先度	86,400，259,200，604,800[秒] (それぞれ 1，3，7 [日])
w	0.5，1，5，10

(ARF) についても試行した．

割り当てルールには，優先度の順番にそのまま割り当てる方法と，積極的 backfill を用いた．

(c) 中断・再開

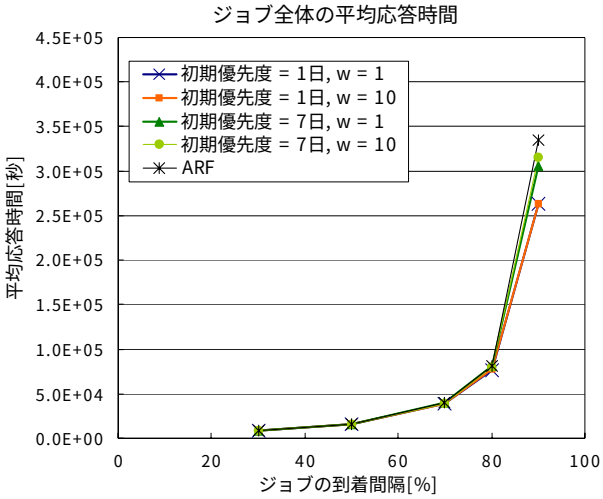
ジョブの中断・再開については，中断・再開がない場合と，中断・再開がある場合についてそれぞれ評価し，中断・再開のコストとして 2，4，8，16 [秒] の 4 通りの値を設定した場合について，それぞれ試行した．

中断は，次のような方法で行う．キュー中で割り当てに充分な優先度を持っているが，必要な計算ノード数が確保できないジョブがある場合，現在ジョブを実行していない計算ノードと，優先度がそのジョブより低いジョブを実行している計算ノードとで必要ノード数が確保できるならば，優先度がより低いジョブを中断する．中断されたジョブは，キュー中に戻され，他の割り当てを待つジョブと同じように扱う．

(d) 統計方法

評価結果では，定常状態中といえるデータを用いる必要があるため，シミュレーションの開始直後と終了直前のそれぞれ総ジョブ数のうち 5%のジョブを対象外とし，10000 個のジョブのうち，501 個目から 9500 個目までの計 9000 個を対象として結果を得た．以下で示す値は，ワークロードに対する 50 回のシミュレーション結果の平均値であり，信頼度 95%の信頼区間は±10%である．

(a) ジョブ全体の平均応答時間



(b) 予約がないジョブの平均応答時間

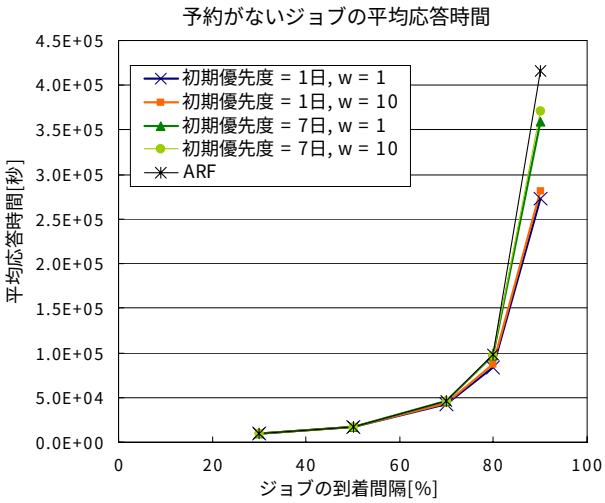


図 2．平均応答時間（中断・再開なし）

4.2. 中断・再開がない場合の結果

本節では，ジョブの中断・再開をしない場合の結果について述べる．ここでは実験結果のうち，予約の割合を 20%，予約の期間を 0-24 時間後，初期優先度を 1 日と 7 日，w を 1 と 10，割り当てルールに積極的 backfill を用いた場合について示す．割り当てルールに積極的 backfill を用いた場合において，予約の割合が 10% のときでは 20% の結果と傾向が同じであった．

図 2 に平均応答時間を示す．ジョブ全体の平均応答時間は，70%以上の負荷において予約に高い優先度を設定することで大きくなり，低い優先度を設定することで小さくなる傾向が見られる．これは予約に優先を置くことにより，予約がないジョブに対しての影響が発生するためである．低

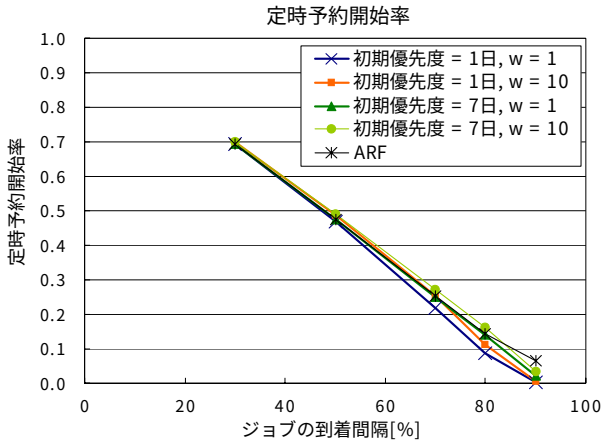


図 3．定時予約開始率（中断・再開なし）

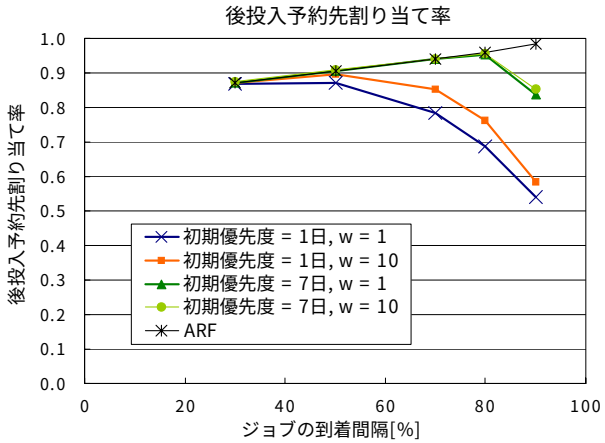


図 4．後投入予約先割り当て率（中断・再開なし）

い負荷の場合において差が見られないのは，キュー中で割り当てを待っているジョブが少ないため，予約の影響が小さいためであるといえる．また，70%以上の負荷では，ジョブ全体と予約がないジョブとの平均応答時間の差が大きくなる．これは高負荷では予約のあるジョブの応答時間が予約がないものと比べ極度に応答時間が小さくなり，ジョブ全体の応答時間が減少するためであると考えられる．

図 3 に定時予約開始率を示す．定時予約開始率は，低負荷から高負荷にかけて下降しており，その値は，1 とジョブの到着間隔との差に近い．これは，ジョブの到着間隔が 30% ならば，予約のあるジョブが割り当て時に 30% の確率で他のジョブを割り当て時に競合をすることになるためであり，1 からジョブの到着間隔の割合を引いた値に近ければ，予約ジョブは予約時間どおりに割り当てられると考えられる．ジョブの到着間隔が 70%以上になると，予約に与えた優先によって下降の速度の差がやや大きくなる．これは，負荷が高くなり，待ち時間が拡大することで，ジョブ間の優先度の差が縮小するため，予約時間に実行が開始できなくなるためである．

図 4 に後投入予約先割り当て率を示す．低負荷から高負荷にかけて上昇し，予約に与える優先が小さいほど，負荷の上昇にともない下落する．低負荷では，他のジョブと割り当てを競合する機会が少なく，高負荷になると競合する機会が高くなるが，予約に対し十分な優先を与えていないと割り当てられることができなくなるためであると考えられる．

以上 3 つの評価指標から，予約がないジョブとジョブ全体の応答時間の差を見ると，定時予約開始率と後投入予約先割り当て率が高いものほど大きくなる．このため，予約が優先されるほど，予約がないジョブが影響を受けるといえる．さらに負荷が大きくなるほど予約を優先するために予約があるジョブが影響を受け，予約の有無での公平性の問題は大きくなるといえる．

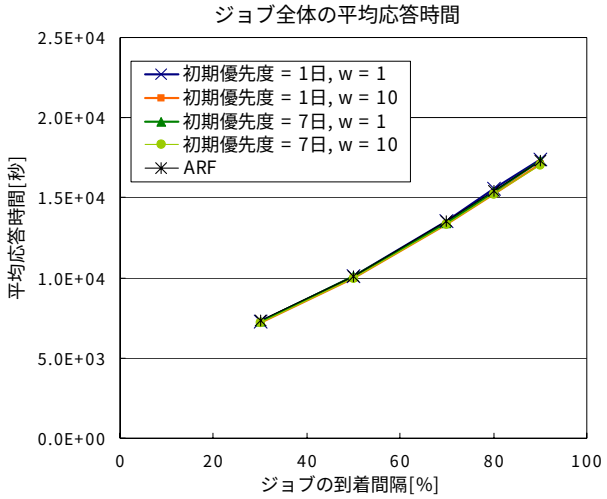
4.3. 中断・再開がある場合の結果

本節では，ジョブの中断・再開をする場合の結果について述べる．ここでは実験結果のうち，予約の割合を 20%，予約の期間を 0-24 時間後，初期優先度を 1 日と 7 日，w を 1 と 10，中断・再開のコストは 8，割り当てルールに積極的 backfill を用いた場合について示す．割り当てルールに積極的 backfill を用いた場合において，予約の割合が 10% のときでは 20% の結果と傾向が同じであり，また中断・再開のコストの差による傾向の変化は見られなかった．

図 5 に平均応答時間を示す．どの場合でも予約の優先による変化は小さい．予約のあるジョブは，他のジョブより高い優先度があれば，実行中のジョブを中断させ，実行開始することができる．予約のあるジョブの応答時間は，負荷の大きさには依存せず，予約がないジョブの応答時間と比較してごく小さい．予約がないジョブは，高負荷になるごとに，割り当ての機会が大幅に減り，待ち時間が長くなり，さらに中断によって応答時間が大きくなる．予約のあるジョブの応答時間は常に小さいため，ジョブ全体の平均応答時間の変化は，予約がないジョブの平均応答時間の変化に従う．

中断・再開がある場合では，充分な優先度があれば，実行中のジョブを中断させて実行開始する

(a)　ジョブ全体の平均応答時間



(b)　予約がないジョブの平均応答時間

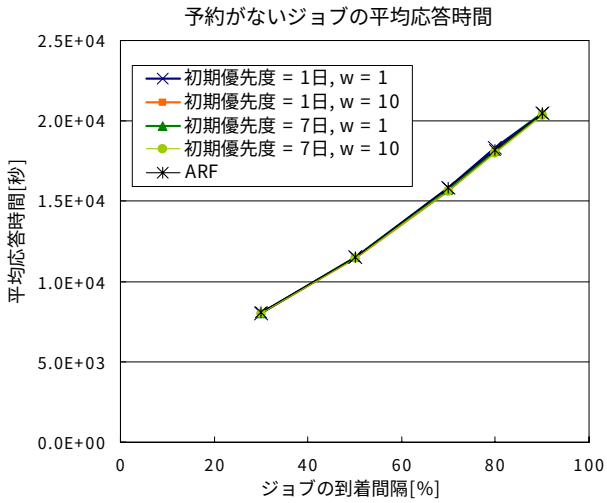


図 5．平均応答時間（中断・再開あり）

ことができるため，定時予約開始率は，ほぼすべてにおいて 0.95 以上の値となり，後投入予約先割り当て率は，すべての結果でほぼ 1 となった．

中断・再開がある場合は，予約に大きな優先をとる必要はなく，予約がないジョブと比較して少しの優先を与えれば，予約が大幅に優先される．そのため，公平性を考えると，予約のあるジョブと予約がないジョブ間で格差が小さくなるようにし，予約がないジョブの割り当ての機会も増やし，予約がないジョブが大幅に悪影響を受けるということを改善する必要がある．

5.　おわりに

本稿では，資源予約機能を持つジョブスケジューリングにおいて，予約のあるジョブと予約がないジョブ間の公平性を考慮したスケジューリング手法を提案するとともに，ジョブ間の公平性を定量的に評価した．

シミュレーションを通して得られた結果では，中断・再開がないときには，負荷が 70%よりも低い場合，予約の影響は小さいと考えられ，負荷が高いときであっても，予約へ予め大きな優先度を与えずに，予約時間とその要求を行った時間との時間差を重視することで高い定時予約開始率が得られた．中断･再開があるときは，結果から，予約へ与える優先を考えるとともに，予約のないジョブに対して割り当ての機会を増加させ中断される機会を減少させる必要があるといえる．

今後の課題としては，様々な割り当ての方法使って割り当てを行うことや，予約のあるジョブの予約時間を保証するために，優先度に応じて予約時間には必ず資源を確保できるようにした場合について調査することが必要である．

参考文献

[1]　藤原 一毅，合田 憲人，"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"，情報処理学会・電子情報通信学会 並列処理シンポジウム JSPP2002, pp.159-160, 2002.

[2]　Warren Smith, Ian Foster and Valerie Taylor, "Scheduling with Advanced Reservations," Proceedings of 14th International Parallel and Distributed Processing Symposium, pp.127-132, 2000.

[3]　Dror G. Feitelson and Larry Rudolph, "Parallel Job Scheduling: Issues and Approaches," Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, Springer-Verlag, vol. 949, pp.1-18, 1995.

[4]　Parallel Workloads Archive. http://www.cs.huji.ac.il/labs/parallel/workload/ .

[5]　Dror G. Feitelson, "Packing schemes for gang scheduling," Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, Springer-Verlag, vol. 1162, pp.89-110, 1996.