2J-6

グリッド環境下での通信遅延を考慮した改良ジョブキュースケジューリング法の評価

岩切　淳一 † 山森　一人 † 吉原　郁夫 † 相川　勝 ††
† 宮崎大学工学部　†† 宮崎大学工学部教育研究支援技術センター

1 はじめに

近年、地理的に離れたサイトにある計算機群をネットワークで接続した計算グリッドに関する研究が盛んに行われている。計算グリッドは複数のサイトから構成され、各々のサイト内には性能の異なる計算ノードが含まれており、それらの負荷状況が刻々と変化する。こうした環境において複数の計算ノードを要求する並列ジョブを効率よく実行するためには、単一の並列ジョブを割り当てるのに必要なサイト数を削減すること、アイドルノード数を削減すること、の 2 点を考慮したスケジューリング法が必要である。

本報告では、並列ジョブを効率的に実行し、グリッド環境のネットワークを考慮したスケジューリング法を提案し、シミュレーションによる評価を行う。

2 スケジューリング手法

2.1 改良ジョブキュースケジューリング Improved Job Queue scheduling(IJQ)

従来のジョブキュースケジューリング法 [1] では、アイドルノード数がキュー先頭のジョブの要求する計算ノード数に満たない場合、ジョブの割り当てが行えず、アイドルノードを有効に利用することができない。そこで、先頭ジョブがアイドルノード数の不足により実行できないとき、後続ジョブを一時的に先行実行するように改良を加えた改良ジョブキュースケジューリング法 (IJQ) を提案した [2]。後続ジョブに割り当てた計算ノードはスケジューラが管理しており、後続ジョブに割り当てた計算ノード数が先頭ジョブの要求する計算ノード数を超えたとき、先頭ジョブに実行権を戻す。その際に後続ジョブはキューに戻されるが、途中結果は保存して再実行に利用するか（マイグレーション）、破棄する。

2.2 サイト数削減割り当て Allocated Sites Reduction(ASR)

地理的に分散したサイトにある計算ノードに対して並列ジョブを割り当てると、広域ネットワークを介することによる通信遅延が無視できなくなる。そこで利用するサイト数をできるだけ少なくなるように計算ノードの割り当てを行い、通信遅延の影響を小さくする。計算ノードの割り当ては個々のサイトが保持するアイドルノード数が多い順に行う。これにより利用するサイト数を少なくする。

3 シミュレーションによる評価

複数のサイトから構成される計算グリッドを想定した環境でシミュレーションを行い、本手法の有効性を示す。シミュレーションを行う際に、マイグレーションにかかるコストと通信遅延コストを以下のように定めた。

3.1 マイグレーションコスト

マイグレーションに必要な時間 $t_{migration}$ は後続ジョブが必要とする計算ノード数 $num_{require}$、グリッド環境の総ノード数を num_{max} とすると、

$$t_{migration} = \frac{num_{require}}{num_{max}} \times penalty,$$

と定める。$penalty$ はマイグレーションにかかる時間に対する重みである

3.2 通信遅延コスト

ジョブを 1 サイト内で実行した時に必要な通信時間を t_{comm_1}、利用するサイト数 num_{site} とした時に必要な通信時間 t_{comm} を、

$$t_{comm} = t_{comm_1} \times \{1 + (num_{site} - 1) \times delay\},$$

と定める。$delay$ は通信遅延に対する重みである。

3.3 評価環境

シミュレーションに用いたパラメータを表 1 に示す。各ジョブは計算時間 t_{calc} と通信時間 t_{comm_1} から構成される。t_{calc} は平均性能の計算ノードで実行したときに必要な計算時間である。t_{comm_1} の平均は各ジョブの総処理時間の 20% を占めるように設定した。この条件下でジョブの平均到着間隔を 200 ～ 300 と変化させ、1000 個の

Evaluation of Improved Job Queue Scheduling Algorithm on Computational Grid with Communication Delay
†Faculty of Engineering, University of Miyazaki
††Technical Center, Faculty of Engineering, University of Miyazaki

表 1　シミュレーション環境

グリッド環境		
サイト数	10	固定
計算ノード数*(num_max)*	1000	固定
計算ノード性能	1000〜2000	一様分布
サイトごとの計算ノード数	30〜100	一様分布
delay	0.1〜1.0	0.1ステップ
ジョブ		
ジョブ数	1000	固定
ジョブが要求する計算ノード数*(num_require)*	1〜1000	一様分布
ジョブに必要な平均計算時間 *(t_calc)*	200	指数分布
ジョブに必要な通信時間 *(t_comm1)*	1〜100	一様分布
ジョブの平均到着間隔	200〜300	ポアソン到着
スケジューラ		
アルゴリズム	Normal/IJQ/ASR/ASR+IJQ IJQ with migration ASR+IJQ with migration	
スケジューリング間隔	10	固定
penalty	100	固定

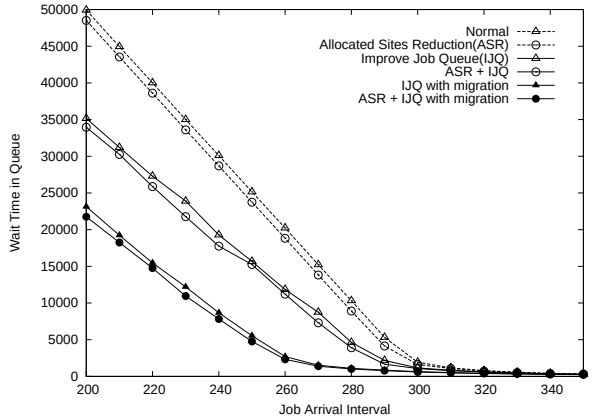

図 1　ジョブ到着間隔に対する待ち時間の変化

ジョブのキューでの平均待ち時間を基に評価を行う。さらに $delay$ を 0.1 ∼ 1.0 に変化させたときも同様に評価を行う。

3.4　評価結果

図 1 に、$delay = 0.3$ に設定したときの、ジョブの到着間隔に対するジョブのキューでの平均待ち時間を示す。図 1 より、IJQ はジョブの到着間隔が短くキューにジョブが溜まりやすい状態であるほど、効果があることが分かる。またマイグレーションを行うと、後続ジョブの一時先行実行が多く行われることで平均待ち時間がさらに少なくなっている。

図 2 はジョブ到着間隔を 300 に設定したときの、通信遅延コストに対するジョブのキューでの平均待ち時間を示している。図 2 より、ASR はサイト間通信を少なくすることでサイト間通信コストが大きい ($delay > 0.3$) とき、ジョブのキューでの平均待ち時間が少なくなっている。さらに ASR は IJQ、マイグレーションと組み合わ

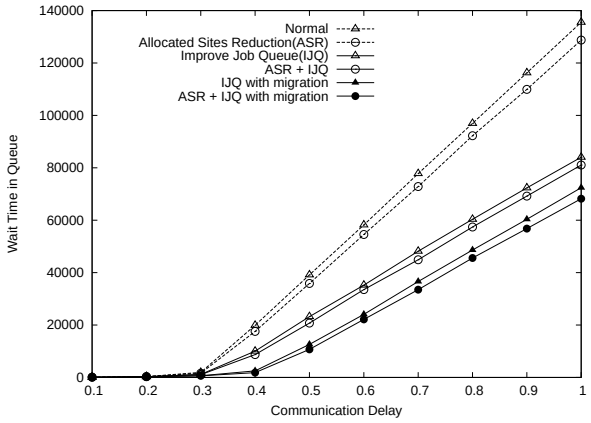

図 2　通信遅延コストに対する待ち時間の変化

せることでキューでの待ち時間をより削減できることが分かった。

4　まとめ

本報告ではグリッド環境でのジョブキューの効率的な利用を可能とする IJQ、通信遅延の影響を小さくする ASR の 2 つのアルゴリズムを提案し、シミュレーションにより評価を行った。IJQ は単体でも従来のジョブキュースケジューリングに比べてジョブのキューでの待ち時間を削減することが出来たが、マイグレーションを行うことで後続ジョブの一時先行実行が促進され、ジョブのキューでの待ち時間をさらに削減できることが分かった。ASR は通信遅延が大きい環境で有効であり、また IJQ やマイグレーションと組み合わせることでより効果を高められることが分かった。

今後はマイグレーションに必要な時間、通信遅延を定義する関数について検討し、本手法の実装による検証、大規模なグリッド環境でのスーパースケジューラと IJQ や ASR の連携手法の検討を行っていく必要がある。

謝辞

本研究の一部は、科学研究費助成金（若手（B）17700239）により行われた。関係各位に感謝する。

参考文献

[1] 譚林、藤本典幸、萩原兼一:グリッド環境における計算ノードの故障を考慮した独立タスクスケジューリングアルゴリズム、情処研報,MPS-043,Vol.2003,No.020(2003)

[2] 岩切淳一、山森一人、吉原郁夫、相川勝:グリッド環境における MPI ジョブを考慮したジョブスケジューリング法の提案、第 58 回電気関係学会九州支部連合大会 (2005)