

概要

近年、グリッドコンピューティングなどのように、ネットワークを用いて多数の計算資源を集めて、大規模な科学技術計算や、Web テキストなどの膨大なデータの処理を行う技術に対する注目が集まっている。このような環境で、複数の計算機が協調して一つの大きなタスクを分割して処理を行うような場合、分割されたタスクの間において、あるタスクで出た結果を次のタスクに利用する、といったようにタスクの間にデータによる依存関係が生じる。そのため、複数の計算機に存在するプロセスの間でデータの共有を行う必要があるが、その場合において用いられるツールの一つが広域分散ファイルシステムである。

広域分散ファイルシステムでは、ある計算ノードでプロセスがファイルを変更した時、それをファイルサーバで反映させ、その後他の計算ノードのプロセスがファイルサーバを参照することでデータのやり取りを実現している。このシステムで問題となるのが複数のプロセスが同時に大きな write を行った場合で、ファイルサーバに負荷が集中し、処理の終了までに時間がかかったり、他プロセスの read の実行を妨げたりする原因となる。しかし、システムの複雑さから、どのプロセスが原因であるか、ということ容易に特定することはできない。そこで、本研究では統計的手法を用いて NFS などの広域分散ファイルシステムに対する負荷の割合を推定する方法の提案、実装を行った。この手法は、分散ファイルシステムの負荷が大きくなった際にどのプロセスが大きな負荷の原因となっているのか、その原因追究の補助をし、分散ファイルシステムを用いた並列分散プログラミングにおけるスケジューリングによるパフォーマンスの改善等へつなげることを目的としている。本稿ではその実装と、部分的な評価について述べる。