

一般計算機によるGrid環境構築技術に関する研究

小板 隆 浩
(Takahiro KOITA)

本研究では、一般計算機からなる Grid 環境でアプリケーションを効率良く実行するための機能について、実装による評価・検討をおこなった。対象アプリケーションは InterProScan とした。InterProScan は、複数のゲノムアプリケーションから構成される統合ゲノムアプリケーションである。

InterProScan は、単一のゲノムアプリケーションからは得られない複数の解析結果を得ることができる。しかし、複数のゲノムアプリケーションを同時に実行するため、たんに質シーケンスの解析には、多くの応答時間が必要である。多数の計算機で InterProScan の処理を適切に分散実行させれば、InterProScan の応答時間を短縮できると考えられる。多数の計算機を利用できる大規模な実行環境に Grid 環境がある。しかし、既存の InterProScan は Grid 環境で実行することができない。既存の InterProScan は、数台程度の計算機がすべてファイル共有されている実行環境を前提としており、小規模な実行環境を対象とした実行方式しか実装されていない。このため、Grid 環境で InterProScan を実行するためには、Grid 環境において、ある程度のファイル共有を実現するファイル管理機能が必要となる。

また、仮に Grid 環境すべての計算機でファイル共有が可能であったとしても、既存の InterProScan のスケジューリング機能では数台程度の計算機しか利用することはできないため、Grid 環境で多数の InterProScan を実行する場合、多くの計算機は有効に利用されず、大幅に応答時間は増加すると考えられる。Grid 環境では、計算機台数は数百台以上となるため、これらの計算機を有効に利用し、処理を適当な計算機に割り当てるスケジューリング機能が必要となる。

本研究では、InterProScan の Grid 環境による応答時間短縮を目的とし、ファイル管理機能とスケジューリング機能を実装した。まず、一般的な Grid 環境では実行することのできない既存の InterProScan を、Grid 環境で実行可能とするために、擬似的なファイル共有を可能とするファイル管理機能を実装した。ファイル管理機能では、実際にファイル共有がない計算機間で InterProScan を実行するために必要なファイルのオンデマンド配置を可能とした。さらに、Grid 環境の多数の計算機を有効に利用し、応答時間を短縮するため、InterProScan の分散実行を対象とした動的なスケジューリング機能を実装した。実 Grid 環境である OBIGrid にて、これらの機能の有効性を比較・検討した。OBIGrid における評価結果より、本研究で実装した各機能の有用性を示した。