様式Ｃ－１９

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成２５年５月１日現在

機関番号：１２６１２
研究種目：基盤研究（C）
研究期間：2010～2012
課題番号：２２５０００４３
研究課題名（和文）　グリッドサーバ自動切替機構の開発

研究課題名（英文）　Research on the Grid-server Auto-switching

研究代表者
本多　弘樹（HONDA HIROKI）
電気通信大学・大学院情報システム学研究科・教授
研究者番号：２０１９９５７４

研究成果の概要（和文）：
計算クリッドにおいて，アプリケーションプログラムの安定した実行を可能とする環境を提供するには，利用可能なサーバの稼働状況に応じてサーバの切り替えを自動的に行う機構が求められる．本研究では，グリッドサーバの自動切替機構の実現を目指して，サーバへのタスク割り当てを行うスケジューリング手法を提案するとともに，サーバ自動切替機能を有するグリッドミドルウェアの開発を行った．

研究成果の概要（英文）：
On a computational grid, to provide the environment where application programs can run stably, the mechanism in which servers to be used are automatically switched according to the execution status of the servers is required. In this research, aiming at realization of the automatic switching mechanism of grid servers, while proposing the scheduling techniques which perform task assignment to servers, the grid middleware which has a grid-server auto-switching function was developed.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2010 年度	2,600,000	780,000	3,380,000
2011 年度	500,000	150,000	650,000
2012 年度	600,000	180,000	780,000
年度			
年度			
総　計	3,700,000	1,110,000	4,810,000

研究分野：総合領域
科研費の分科・細目：情報学・計算機システム・ネットワーク
キーワード：ハイパフォーマンスコンピューティング，並列処理，グリッド

１．研究開始当初の背景

広域ネットワーク上のコンピュータ群を統合し単一システムとして利用可能とする計算グリッドは，並列処理による高い計算性能を備えており，高性能計算環境として魅力的なものである．しかし，計算グリッドでは，その規模の大きさや構成機器の非均質性から，その利用にあたり煩雑な作業が必要で，必ずしも一般ユーザに使いやすいものとなっていなかった．そこで，これら煩雑な作業を軽減または自動化し，計算グリッドの簡便な利用環境を提供するシステムの研究が国内外で進められていた．

このような中で応募者は，グリッド利用における作業のうち，遠隔サーバへのライブラリの配置やグリッドサーバの切り替えを自動化するための研究課題に取り組んできていた．

具体的には，サーバ切り替えの自動化に必要な機能とそれを統合する方法を検討し，サーバ稼動状況監視を自動的に行い，監視結果に連携して必要に応じて自動的にサーバ切り替えを実行するという枠組を提案してきた．さらに，パイロットシステムの構築を通して，サーバ稼働状況の自動監視機能と代替サーバへの自動切り換えに必要なサーバの動的管理機能の連携という本枠組の基本部分が実現可能であることを実証するに至っていた[1]．

２．研究の目的

アプリケーションプログラムをグリッド上で効率良く実行させるには，グリッドサーバの稼働状況の変化に対応して，サーバの適切な切り替えが必要となる．サーバ切り替えを行うには，サーバの稼動状況の監視，適切な代替サーバの選択，サーバへのタスクのスケジューリング，サーバへのアプリケーションやライブラリの配置などの作業が必要となる．

本研究課題では，計算グリッドを簡便に利用するための環境の構築を目指し，アプリケーション実行時におけるサーバの適切な動的切り替えを自動化するシステムを構築するために必要な機能を明らかにするとともに，これを実現するミドルウェアを開発し，その有用性を明らかにすることを目的とする．ミドルウェア開発に際しては，その利用にあたりできるだけソフトウェアのインストール等に手間のかからない方法を採用することとする．

３．研究の方法

研究は３年間かけて行うこととした．まず，これまで提案してきた自動的なサーバ切り換えの枠組みに基づいて必要機能の検討を行い，本研究で注力すべき点を明らかにする．次に，必要な各種機能を備えたグリッドミドルウェアを実装しその評価を行う．

具体的には次のとおりである．

(1) 自動的なサーバ切り換えの枠組を構成する各種機能として，従前のものに加えて必要となる機能を精査し，実現可能なものであるかを検討する．
(2) 各種機能について，既存のグリッドシステム，例えばGolbus Toolkit，Inca，G-Monitor，Ninf-Gなどに装備されている機能を利用できる否かを検討し，既存のものを利用するものと新規に開発すべきものを峻別する．
(3) 動作検証に用いるグリッドテストベッドを構築する．
(4) 自動的なサーバ切り換えの各種機能のうち新規に開発すべきモジュールの実装を行う．
(5) 実装に際しては，当初は本研究のシステムでスタンドアロンなものとするが，他の研究機関での開発状況に応じて，他のシステムとの連携を取れるよう柔軟に対処する．
(6) 実装した各機能を統合し，ひとつのグリッドミドルウェアとして構築する．グリッドミドルウェアはできるだけインストール等の手間のかからないもととして構築することを目指す．
(7) 構築したグリッドテストベッド上で開発したグリッドミドルウェアの動作検証と有効性の評価を行う．
(8) 研究成果を学会等で発表するとともに研究成果をまとめる．
(9) 研究体制は，研究代表者(本多弘樹)と１名の研究協力者(各年に大学院学生１名程度)で構成する．

４．研究成果

本研究の主要な成果は，ローカルサイト内でのサーバ切替のためのローカルスケジューリング手法の開発，ローカルサイト間でのサーバ選択のためのグローバルスケジューリング手法の開発，サーバ自動切替機構を有するWebベースグリッドシステムの開発である．

それぞれの研究成果の概要は次のとおりである．

(1)ローカルサイト内でのサーバ切替のためのローカルスケジューリング手法

グリッド環境でのサーバ切替のための機能を検討の結果，ローカルサイト内でサーバへタスクを割り当てるスケジューリング機能の重要性が明らかになった．

そこで，グリッド環境中のローカルサイト内の構成要素としてのヘテロジニアス計算機クラスタにおいて，グローバルスケジューラによってローカルサイトに割り当てられたタスクをローカルサイトのサーバ群にスケジューリングするための手法の開発を行った．

開発した手法の特徴は，いったん仮スケジュールされたタスクの割当先サーバを変更しタスクの再割り当てを行い，より低消費電力なサーバでそのタスクを実行することにより，近年問題となっているサーバによる消費エネルギーを低減することを可能とする点にある．

具体的には，提案スケジューリング手法は，処理時間のみに着目した既存のスケジューリング手法 Heterogeneous Earliest Finish Time(HEFT)法[2]を用いて仮スケジューリングを行い，そのスケジュール結果をもとに，サーバのアイドル時またはスタンバイモード時の消費電力を考慮しつつ，タスクセット

の処理時間の増大を抑えながらサーバへのタスクの再割り当てを行うものである.

タスクの再割り当ては次の方法で行う.

(ア) 再割り当ての対象とするタスクは,タスクセットのタスクグラフ上でクリティカルパス上にないタスクとする.

(イ) 再割り当て先のサーバは,再割り当て対象タスクの最早処理開始可能時刻から最遅処理完了可能時刻までの間に,当該タスクの処理時間より長い時間他のタスクが割り当たっていないサーバを候補サーバとし,その候補サーバのうちタスクセット全体の消費エネルギーが最も小さくなるサーバとする.

提案手法の評価に際しては,提案手法と従来の HEFT 法のそれぞれのスケジュール結果について,タスクセットの処理時間と消費エネルギーに関して比較を行った.

一つのタスクセット中のタスク数は 100 個とし,ランダムタスクセット 180 例を評価対象とした.サーバ数を 12,24,36 と変化させ,タスク処理時間の最大最少比が 1.5 と 2.5 の場合,消費電力の最大最少比が 1.5 と 2.5 の場合について実験を行った結果,HEFT 法に対して提案手法は最大で 13.4%の消費エネルギー削減が達成できることが示された.

また,提案手法は HEFT 法に比べスケジューリングのための処理時間が長くなることが懸念されたが,表1に示す通り,スケジューリングのための処理時間は長くはなるものの許容範囲内であることが確認された

表1 スケジューリング時間(秒)

ｻｰﾊﾞ数	12	24	36
HEFT 法	0.32	0.60	0.94
提案手法	1.44	3.19	5.74

この成果は,グリッド上でアプリケーションを実行する際,低消費電力で実行できるようにサーバ切替を可能とする点で重要な技術である.

本研究成果の一部は,情報処理学会の計算機アーキテクチャ研究会で発表を行った(学会発表の(1)).

(2)ローカルサイト間でのサーバ選択のためのグローバルスケジューリング手法

複数のローカルサイトで構成されるグリッド環境において,ローカルサイトで発生したタスク群をより負荷の軽いリモートサイトのサーバ群に割り当て,ローカルサイト間での負荷を均等化しグリッド環境全体で効率良い実行を可能とするスケジューリング機能を検討した結果,グリッド規模が増大するとともにグローバルスケジューラの負荷が高くなり,スケジューラ自体がボトルネックとなる問題点に着目した.

そこで,ローカルサイト間での負荷を均等化し,グリッド環境全体で効率良い実行を可能とするグローバルスケジューリングのための手法の開発を行った.

開発したスケジューリング手法は,ローカルサイト間でジョブ割り当てを行う際にボトルネックとなりうる集中管理スケジューラを用いず,スケジューラがボトルネックとなることを回避するために,各ローカルサイトが分散してセルフスケジューリングを行う方式を採用した.すなわち,各サイトのスケジューラは自サイト内のサーバ資源だけを管理し,サイトスケジューラ間を相互通信によって連携させることで,自サイトより低負荷なサイトを探索し,そのサイトにタスクを割り当てるものである.

開発した手法の特徴は,分散されたグローバルスケジューリングに際して,サイトの負荷の見積方法,サイトの決定方法,サイトを探索する際の開始サイトの決定方法を工夫した点にある.

負荷の見積には,従来用いられていたローカルサイトスケジューラキュー内のジョブ数だけではなく,各ジョブの予想実行時間の総和をも用いることとした.

サイトの決定方法としては,他サイトにキュー内での待ち時間を順次問い合わせ,自サイトより待ち時間が短いサイトが見つかった時点でそこにジョブを割り当てる FirstFit(FF)方式とあらかじめ設定された探索範囲内のすべてのサイトにキュー内での待ち時間を問い合わせ,待ち時間が一番短いサイトを選択しジョブを割り当てる BestFit(BF)方式を考案した.

探索開始サイトの決定方法としては,すべてのサイトを区域に分割し,その区域の中からスケジューラのキュー内での待ち時間の平均が最も小さい区域を選択する BestArea(BA)方式,自サイトが属する区域のスケジューラのキュー内での待ち時間の平均値が短い区域の中からランダムに区域を選択する Random Better Area(RBA)方式,および,区域の分割はせずにすべてのサイトの中からランダムにサイトを抽出して,自サイトを含めキュー内での待ち時間が短いサイトを選択する Start Node Random(SNR)方式を提案した.

評価に際しては,提案手法を各種組み合わせたスケジューリング手法について,GridSim[3]を用いてスケジュール長のシミュレーションを行った.

その結果,ジョブの発生サイトが分散している場合には,サイト探索範囲を限定し BF 法と RBA 法を組み合わせた方法が負荷バランスとジョブの待ち時間の短縮効果が最も高く,ジョブの平均待ち時間は従来手法に比べて約 38%短縮できることが確認できた.また,

ジョブの発生サイトに偏りがある場合には，サイト探索範囲を限定し BF 法と SNR 法を組み合わせた方法が負荷バランスとジョブの待ち時間の短縮効果が最も高く，ジョブの平均待ち時間は従来手法に比べて約 44%短縮できることが確認できた．

この成果は，サーバへのジョブの初期割当，および，サーバ切替の際のジョブの再割り当ての際に，サーバの負荷均等化をはかり，グリッドシステム全体での処理効率を向上させるために重要な技術である．

(3)サーバ自動切替機構を有する Web ベースグリッドシステム

本研究の初期の検討段階で，グリッドシステムの利用にあたって，ミドルウェアのインストールの手間を少なくすることが必要であることを強く認識した．

そこで，グリッド環境において，ローカルサイトのクライアントで発生したタスクをリモートサイトのサーバ群に割り当てる際に必要となる仕組みとして，特殊なミドルウェアのインストールなしに実現できる手法を検討し，開発を行った．

本手法の特徴は，ジョブを投入する利用者，および，サーバを提供する利用者は，新たなミドルウェアやプラグイン等のインストールを一切必要とせずに，既存の Web ブラウザを使用するだけで，Web サーバとの連携により，ジョブの投入，実行，実行結果の提示を行えるようにしている点にある．サーバの自動切替に必要となるサーバの参加および脱退の検知は，ハートビート機能の実装により実現している．

提案するグリッドシステムの具体的な構成を図１に示す．

図中， Web サーバは，クライアントとサーバ用の Web ページを提供するとともに，ジョブ管理機能，サーバ管機能などの統合的な機能を有する．クライアントでは，ユーザが Web ブラウザでクライアント用 Web ページをアクセスすることにより，ジョブの投入とジョブ実行結果取得を行う．サーバでは， ユーザが Web ブラウザでサーバ用 Web ページをアクセスすることにより，サーバ登録・脱退を行う．サーバでのジョブやデータのダウンロード/アップロード，ジョブの実行，ハートビート送出は自動的に行われる．

サーバにおいては，Web ブラウザ上でジョブを実行することにより高負荷がかかり Web ブラウザの操作に影響が出ることを抑制するために，HTML5 の機能である Web Workers を利用して，ジョブ実行をバックグラウンドで行うよう実装している．

Web サーバでは，ハートビート検知機構によりサーバからのハートビートを監視し，ハートビートが途切れ設定されたタイムアウト時間が経過した際には，そのサーバが利用不可能となったと判断し，スケジューリング機構へサーバの切替と新しいサーバへジョブの再割り当てを行うよう通知する．

本システムの有用性・実用性を検証するため，テストベッドシステム上に本システムを実装し評価実験を行った．

その結果，サーバの動的な参加・脱退に追従して，サーバの自動切替が正しく行われ，必要に応じてジョブの再割り当てを行い安定したジョブ実行が行われていることが確認できた．

また，サーバが増加した際に Web サーバがボトルネックとなり，Web サーバでのハートビート検出処理の遅延が懸念されたが，実験の結果サーバが 50 台程度までなら問題なくハートビート検出処理できることが確認された(表 2)．

表 2 ハートビート(HB)処理実験結果

サーバ数	10	20	30	40	50
処理時間[ms]	1510	1511	1514	1510	1513
HB 処理数[/s]	0.662	0.661	0.660	0.662	0.660

(ハートビート間隔 1[s]，タイムアウト 7[s])

この成果は，自動的なサーバ切替を行うグリッドシステムを，ミドルウェア等のインストールなしに実現するために重要な技術である．

なお，本研究課題の前述の（１）～（３）の研究に関連して３名の博士前期課程学生の研究指導を行い，それぞれの学生は，「コンピュータクラスタにおける省エネルギー

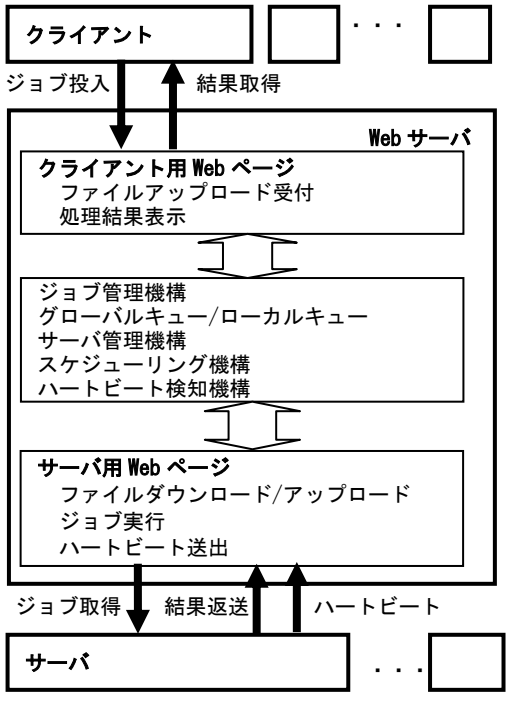

図１　提案グリッドシステムの構成

化のためのスケジューリング手法の研究」,「大規模なグリッド環境におけるジョブスケジューリング手法に関する研究」,「グリッド環境の動的な変化を考慮したWebベースPCグリッドシステムの研究」と題する学位論文を作成し修士学位を取得した.

＜参考文献＞
[1] H. Watanabe, S. Hirasawa, H. Honda “F-Omega: a Framework for Steering GridRPC Applications”, Third IEEE International Conference on e-Science and Grid Computing, pp.475-482, Dec. 2007.
[2] H.Topcuuouglu et.al “Performance Effective and Low Complexity Task Scheduling for Heterogeneous Computing”, TEEE Trans. Parallel Dist. Systems, Vol.10, No8, pp.795-812(1999).
[3] B. Buyya et.al “GridSim: A Toolkit for the Modeling and Simulation of Distributed Resource Management and Scheduling for Grid Computing”, The Journal of Concurrency and Bomputation: Practice and Experince, Vol 14, Issue 13-15, Wiley Press, pp.126-131, 2008.

５．主な発表論文等
（研究代表者，研究分担者及び連携研究者には下線）

〔雑誌論文〕（計０件）

〔学会発表〕（計１件）
(1) 山下良,近藤正章,平澤将一,本多弘樹:ヘテロジニアス計算機クラスタにおける省エネルギー化タスクスケジューリング手法，情報処理学会　計算機アーキテクチャ研究会ARC-194(3),2011年3月10日，高知県香美市.

〔図書〕（計０件）

〔産業財産権〕
○出願状況（計０件）

○取得状況（計０件）

〔その他〕
６．研究組織
(1)研究代表者
本多 弘樹（HONDA・HIROKI）
電気通信大学・大学院情報システム学研究科・教授
研究者番号：２０１９９５７４