RCSS ディスカッションペーパーシリーズ ISSN-1347-636X
第 89 号　2009 年 10 月
Discussion Paper Series
No.89 October, 2009

インターネットを介した PC グリッドシステムの構想

榎原博之・梅本潤志・森川浩明

RCSS

文部科学大臣認定　共同利用・共同研究拠点
関西大学ソシオネットワーク戦略研究機構
関西大学ソシオネットワーク戦略研究センター
（文部科学省私立大学学術フロンティア推進拠点）

Research Center of Socionetwork Strategies,
“Academic Frontier” Project for Private Universities, 2003-2009
Supported by Ministry of Education, Culture, Sports, Science and Technology
The Research Institute for Socionetwork Strategies,
Joint Usage / Research Center, MEXT, Japan
Kansai University
Suita, Osaka, 564-8680 Japan
URL: http://www.rcss.kansai-u.ac.jp
http://www.kansai-u.ac.jp/riss/index.html
e-mail: rcss@ml.kandai.jp
tel: 06-6368-1228
fax. 06-6330-3304

インターネットを介したPCグリッドシステムの構想

榎原 博之* 梅本 潤志† 森川 浩明‡

2009年10月

概 要

PCの性能向上はめざましく、複数台のPCをネットワークに接続して大規模な計算を処理するPCグリッドシステムが注目されてきている。本研究では、PCクラスタをインターネット（WAN）上に分散配置した分散PCグリッドシステムの構想を提案する。本構想では、各PCクラスタには管理サーバとしてグリッドエージェントを配置し、グリッドエージェント間で通信を行うことにより、効率の良い計算資源配分と通信負荷の軽減を実現する。さらに、グリッドエージェントにはグリッドポータルの機能を持たせ、各地点で投入した計算ジョブは地点に関係なく効率良く分散実行できるシステムの実現を目指す。

Keyword: グリッド，PCクラスタ，仮想化，グリッドエージェント，分散配置

*関西大学 システム理工学部, 関西大学 ソシオネットワーク戦略研究センター研究員, Email: ebara@kansai-u.ac.jp
†関西大学 工学部 電子情報システム工学科, Email: thx.thx@hotmail.co.jp
‡大阪市立大学 創造都市研究科, Email: paildriver@gmail.com

Conception of Distributed PC Grid System on the Internet

Hiroyuki Ebara[1], Junji Umemoto[2], and Hiroaki Morikawa[3]

October, 2009

Abstract

The improvement in performance of PC is remarkable, and PC grid system, which connects a number of PCs to a network and performs a large-scale computing job, has attracted attention. In this research, the concept for the distributed PC grid system consisting of distributed PC clusters on the Internet (WAN) is proposed. With this concept, efficient computing resource allocation and mitigation of communication load are realized, by seting a grid agent as a management server to each PC cluster and communicating among the grid agents. Furthermore, a grid agent has the function of a grid portal, and we aims at realization of the system which can execute a computing job efficiently on any PC cluster regardless of a PC cluster which it is submitted to.

Keyword: Grid Computing, PC Cluster, Vertualization, Grid Agent, Distributed Allocation

[1]Faculty of Engineering Science, Kansai University, Research Fellow, The Research Center of Socionetwork Strategies, Email: ebara@kansai-u.ac.jp

[2]Department of Electronic, Faculty of Engineering, Kansai University, Email: thx.thx@hotmail.co.jp

[3]Graduate School for Creative Cities, Osaka City University, Email: paildriver@gmail.com

1　はじめに

コンピュータのめざましい発展に伴い、家庭や会社で使われている PC でさえ高度な科学技術計算が可能となっている。一説には、現在の PC の性能は初代スーパーコンピュータの性能の 100 倍以上といわれている。これらの PC を複数台使えば、ある程度の規模の高速計算が可能であると思われる。さらに、それらの PC は 24 時間稼働しているわけではなく、それらの PC の遊休時間を利用すれば、安価で高速計算が可能となる。したがって、複数の PC を LAN でつないだ PC クラスタシステムや PC グリッドシステムのように安価で計算パワーを供給できるシステムの開発は重要であり、コンピュータメーカを中心に開発が進められている。特に、PC グリッドシステムは、高性能な PC がその性能をフルに発揮していない点を考慮し、PC の有効活用を目的として、手軽に計算パワーを得ることができるシステムを目指している。

グリッドシステムの定義は、元来、電力系パワーグリッドに例えられるようにネットワークに接続するだけで、誰もが安価にあるいは無料で計算パワーを利用できるシステムである [4][1]。しかし、実際に開発されているグリッドシステムは、専用機にグリッドコンピューティング用ソフトウェア（ミドルウェア）を導入した PC クラスタシステムの延長線上のものをインターネットなどの広域ネットワーク（WAN）に接続したもので、一般ユーザが計算パワーを自由に使えるものとは到底言い難い。また、広義のグリッドシステムに属する PC グリッドシステムは、家庭などで使われている PC を使って大規模計算を行うシステムと定義されている。しかし、現在のところ、SETI@home プロジェクト [11] に代表されるように、PC 側がサーバにあるソフトウェアをダウンロードしてバックグラウンドで実行するしくみで、PC 起動中に遊休状態になったとき、そのソフトウェアがサーバからデータを取得し、計算し、実行結果をサーバに返すしくみとなっている。このため、独立した小問題に分割できる問題にしか適用できず、プロセス間で通信を行う並列計算には向かない。また、停止中の PC は、計算ジョブを実行できない。

著者らは、停止中や遊休の PC をグリッドシステムの計算サーバとして利用できるシステムを開発した [2] [9]。このシステムは、大学内のコンピュータ演習室などにある PC の有効利用を想定しており、グリッドサーバが休止している PC を見つけ、ネットワークを介して起動させる。さらに、マイグレーション機能によって演習室内でユーザが利用を開始すると計算を行っている仮想マシンをサスペンドさせ、計算内容を他の PC に移行させる機能を持っている。この機能により、長時間ジョブの実行が可能である。

一方、VMware[13] や Xen[14] に代表される仮想マシンソフトが注目されている [7][8]。仮想マシンソフトを使うことにより、1 台のサーバで複数の OS を稼働させ、複数のサーバの機能を 1 台で賄って、サーバの有効利用を可能にしている。また、仮想マシン上の OS（ゲスト OS）はハードウェアと独立しており、仮想ディスクイメージを転送して別の PC 上で復元すること（マイグレーション）が可能である。マイグレーションは、主に、サーバの故障やメインテナンス時の対応、さらには、サーバの負荷分散に利用されている。

本研究では、PC クラスタをインターネット（WAN）上に分散配置した分散 PC グリッドシステムの構想を提案する。本構想では、各 PC クラスタには管理サーバとしてグリッドエージェント

を配置し、グリッドエージェント間で通信を行うことにより、効率の良い計算資源配分と通信負荷の軽減を実現する。さらに、グリッドエージェントにはグリッドポータルの機能を持たせ、各地点で投入した計算ジョブは地点に関係なく効率良く分散実行できるシステムの実現を目指す。

2　グリッドシステム

2.1　グリッドシステムとは

グリッド（Grid）とは、辞書を引くと「格子、碁盤の目、送電網（Power Grid）」などと記述されている。グリッドシステムにおけるグリッドは、電力システムで用いられる送電網に由来する。広義の意味でのグリッドシステム（Grid System）とは、「コンセントに差し込めばいつでもどこでも必要なだけ電力が得られるように、情報コンセントに接続するだけで、コンピュータネットワークを介して、誰もが安全に（強固なセキュリティ）、安定して（必要な時に必要なだけ提供可能）、安易に（システム側を気にすること無く利用可能）、ネットワーク上のコンピューティング資源を利用できるシステム」と定義することができる。

グリッド協議会のホームページ [5] によると、グリッドシステムの定義（狭義）は、「グリッドは、広域ネットワーク上の計算、データ、実験装置、センサー、人間などの資源を仮想化・統合し、必要に応じて仮想計算機（Virtual Computer）や仮想組織（Virtual Organization）を動的に形成するためのインフラ」と記されている。

2.2　グリッドシステムの機能

グリッドシステムの分野で第一人者である、米シカゴ大学兼アルゴンヌ国立研究所の Ian Foster 氏は、以下の三つの機能をグリッドのチェックリストとして挙げている [3]。

1. coordinates resources that are not subject to centralized control …
   集中管理されていない分散した資源のコーディネート
2. … using standard, open, general-purpose protocols and interfaces …
   オープンスタンダードなプロトコルやインターフェースの利用
3. … to deliver nontrivial qualities of service.
   単純には得られない質の高いサービスの提供

2.3　グリッドシステムの種類

広義のグリッドシステムには、データグリッド、コンピューティンググリッド、ビジネスグリッド、キャンパスグリッド、PC グリッドなどがある。

- データグリッド
  大規模なデータを処理するシステム

- コンピューティンググリッド
  高度な計算を高速に処理するシステム
- ビジネスグリッド
  企業内情報システムを統合化するための信頼性の高いウェブベースのシステム
- キャンパスグリッド
  大学キャンパス内の PC を有効活用して科学技術計算を処理するシステム
- PC グリッド
  家庭などにある遊休 PC を集めて大規模な分散計算を行うシステム

2.1 節で述べた狭義のグリッドシステムの定義では、PC グリッドを含めず、柔軟な情報サービスの提供を目的としたシステムを指す。広義では、PC グリッドのような遊休 PC を活用する分散計算システムまで含める。

2.4 PC グリッドシステム

PC グリッドシステムとは、前項でも述べたように遊休 PC を有効活用するシステムである。最も有名かつ最大のプロジェクトは、SETI@HOME プロジェクトである [11]。このプロジェクトは、地球外知的生命体探査（SETI）を行うもので、プエルトリコの天文台で収集された宇宙から届く電波を解析し、人工的に発信されたと思われる電波を検出するプロジェクトである。参加希望のボランティアは、SETI@HOME のホームページからソフトウェアをダウンロードし、PC にインストールすれば、簡単に参加できる。このソフトウェアは、主にスクリーンセーバーのように PC を利用していないときにデータを取り込み、解析を行う。現在、多くのボランティアを得、トップクラスのスーパーコンピュータに匹敵するデータ処理を実現している。

PC グリッドシステムは上記の SETI 型グリッドシステムが一般的である。すなわち、各 PC が自発的に計算処理に参加し、グリッドサーバは単にジョブの投入と結果の収集のみを行う。SETI 型 PC グリッドシステムの研究として、中部電力の曽山らの研究がある [12]。著者らのシステムでは、グリッドサーバが能動的に遊休 PC を探索し、遊休 PC にジョブを投入する PC グリッドシステムについて研究している [2] [9]。

2.5 PC クラスタシステム

複数の PC を利用するシステムとして PC クラスタシステムがある。PC クラスタシステムは、複数の PC を LAN などのネットワークに接続し、一つの並列計算機システムとして機能させるもので、中にはスーパーコンピュータに匹敵する性能を持ったものも存在する。PC クラスタシステムでの PC は、一般には計算専用で日常業務には利用しない。PC グリッドシステムとの最も大きな違いは、LAN などのローカルネットワーク内でのシステムで、インターネットを介したシステムは想定していない。図 1 に示すように、PC クラスタシステムは、グリッドシステムの 1 拠点と考えることができる。

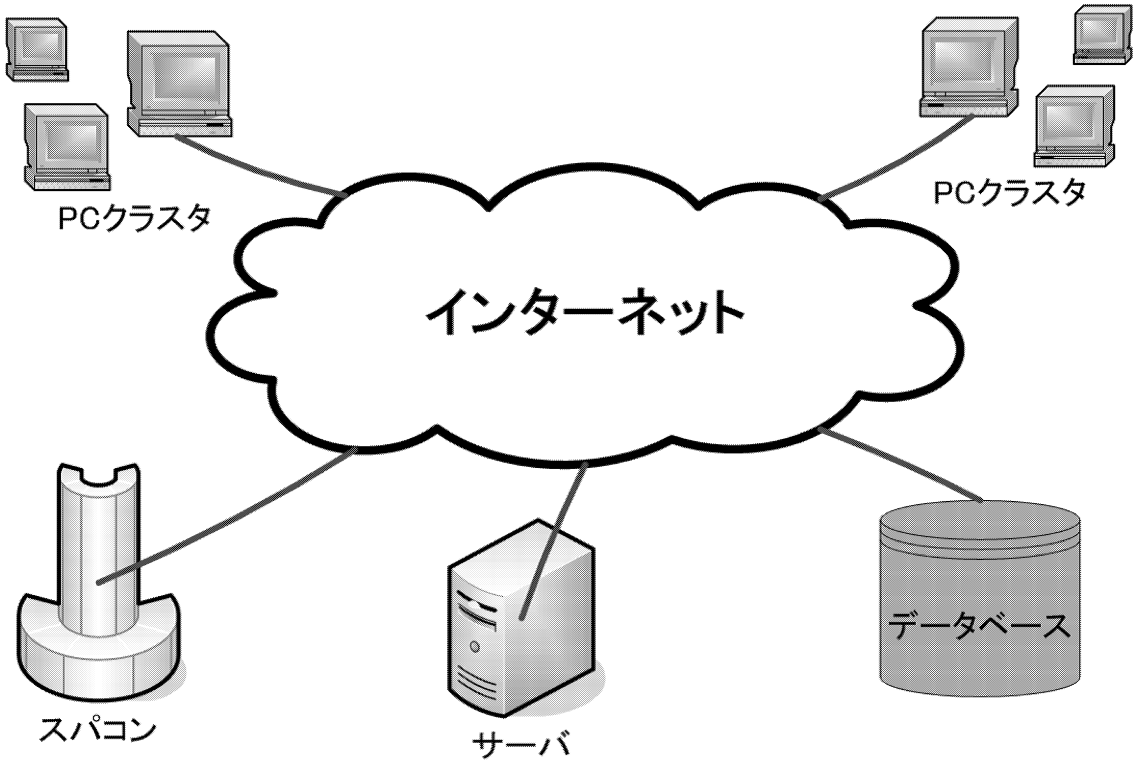

図 1: グリッドシステム

3　仮想マシン

3.1　仮想マシンとは

仮想マシンとは、コンピュータ上に CPU やメモリ、通信回線などを仮想的に構築し、コンピュータ内に複数の仮想的なコンピュータを構築する技術である。代表的な仮想マシンソフトとしては、VMware[13] や Xen[14]、Hyper-V[6]、Parallels Desktop[10] などがある。仮想マシンは、ハードウェア上に仮想化層を構築し、仮想化層を通して間接的にハードウエアを操作している（図 2)。

仮想マシンの一般的な利用法は、

- 1 台の PC に複数の OS を起動させる
- ハードウェアに依存せずにシステム開発を行う
- 1 台の PC に複数のサーバを立ち上げ、有効利用を図る

などが挙げられる。仮想マシンの利点は、ハードウェアに依存しないシステムを組めることである。

3.2　マイグレーション

代表的な仮想マシンの機能の一つにマイグレーションがある。マイグレーションとは、サーバ内のシステムやデータを他のサーバに移行させることである。一定期間ごとに現在実行中の状態を保存し、障害発生時などに別の PC で保存した状態を展開するチェックポイントマイグレーショ

図 2: 仮想マシン

ンや、サーバを停止させることなく、移行させることができるライブマイグレーションなどがある。このように、ハードウェアに依存しない仮想マシンでは、ハードディスクのイメージ（プログラムやデータ）をそのまま移行させることが可能となっている。

4　分散 PC グリッドシステム

4.1　分散 PC グリッドシステムの概略

著者らが提案する分散 PC グリッドシステムは、インターネット（WAN）上に分散されている PC クラスタをつなぎ、それらを一つのシステムとして機能するものである。PC クラスタ内の各 PC の状態を把握したり、PC クラスタ間の情報交換のために各 PC クラスタにグリッドエージェントと呼ぶ管理サーバを設置する。グリッドエージェントはグリッドポータルの機能も有し、各地点で計算ジョブの投入が可能である。

具体的には、図 3 に示すように、関西大学システム理工学部のアルゴリズム工学研究室（著者らの研究室）と、関西大学ソシオネットワーク戦略研究センター（RCSS）、大阪市立大学学術情報総合センターの 3 地点に PC クラスタとそれを管理するグリッドエージェントを設置し、グリッドエージェント間で通信することにより、効率の良い計算資源配分を実現する。

ユーザは、各地点のグリッドエージェント（グリッドポータル）から計算ジョブを投入する。投入された計算ジョブは、グリッドエージェント間の情報交換により、最も適した PC クラスタで実行される。投入された地点の PC クラスタで実行されるとは限らない。さらに、一つの PC クラスタでの実行が難しい場合は、複数の PC クラスタをまたがった実行も可能である。この実行は、プロセス間通信の必要が無いパラメータスイープ型のジョブだけでなく、プロセス間通信が必要なジョブでも実行できる。ただし、複数の地点での実行は、通信遅延などの問題が起こる可能性

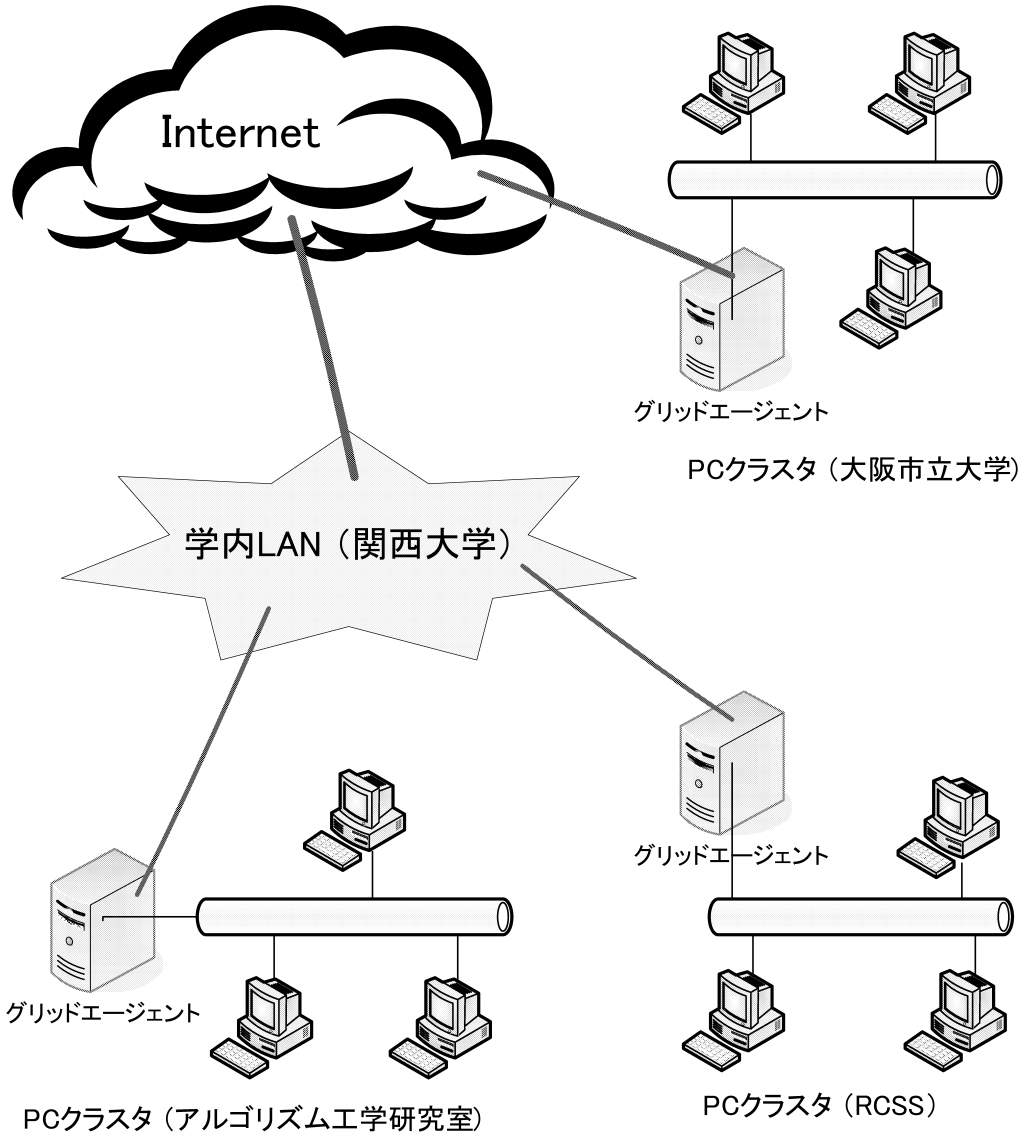

図 3: 分散 PC グリッドシステム

があるため、ユーザ側であらかじめ、PC クラスタを分けることができる範囲を指定し、PC クラスタ内通信と PC クラスタ間通信を区分する必要がある。

これらを実現するために、グリッドエージェントは以下のような機能を備えている。

- 各 PC のコア単位での負荷状況の把握
- 計算ジョブの転送
- プロセス間通信の仲介
- マイグレーションされるジョブイメージの転送
- グリッドポータル機能

4.2 分散 PC グリッドシステムの特徴

提案する分散 PC グリッドシステムは、PC クラスタ間でのプロセス間通信を実現することにより、PC クラスタをまたがった効率の良い計算資源配分が可能となる。さらに、PC の CPU がマルチコア化していることから、マルチコアを考慮した計算資源配分を行う。具体的には、コア単位での PC の負荷状況を把握し、コア単位での計算ジョブの投入や PC 単位でのコア分散を考慮した並列計算ジョブの投入を可能とする。また、プロセス間通信は、暗号化によりセキュリティを保証する。

本システムは、仮想マシン上で構築するため、計算ジョブのマイグレーションが可能である。それにより、ジョブ実行中の PC を途中で中断させ、他のジョブを実行させたり、ユーザが単に PC として利用したりすることが可能である。この機能により、大学などの計算機室の PC をオープン利用時でも計算サーバとして利用することが可能となる。結果として、長時間のジョブも実行可能である。

提案する分散 PC グリッドシステムの特徴を以下に列挙する。

- 効率の良い計算資源配分
- マルチコアを考慮したジョブ投入
- 暗号化によるグリッドエージェント間通信
- 計算ジョブのマイグレーション
- ジョブ実行中 PC の途中中断
- マイグレーションによる長時間ジョブの実行
- 休止中 PC の電源投入
- グリッドエージェント間は対等
- ジョブイメージの効率の良い転送
- 分散配置されたグリッドポータル

4.3 分散 PC グリッドシステムの開発

提案する分散 PC グリッドシステムの構成は、複数のグリッドエージェントがインターネット上に分散配置されており、各グリッドエージェントは Ethernet Switch を介して PC クラスタを構成している。グリッドエージェントと各 PC の OS には Linux を採用する。

1. PC クラスタ

   管理サーバであるグリッドエージェントでは、クラスタ内の各 PC の計算資源やジョブの実行をコア単位で管理する。グリッドエージェントと各 PC には OpenMP や、MPI、MPC++ などの並列計算ライブラリをインストールし、並列コンピューティング環境を構築する。グリッドエージェント間では、並列計算ライブラリから提供される P2P 方式の通信でグリッドエージェント同士が対等で密な通信を行う。さらに、コア単位でのジョブ実行だけでなく、

PC 単位でのジョブ実行も可能であり、CPU 内のコアを効率良く処理するジョブの実行も可能である。

2. 計算ジョブのマイグレーション

   各 PC は仮想マシンソフトである Xen を利用し、準仮想化で仮想マシンモニタと仮想マシンの二つを動作させる。Xen の準仮想化技術を利用することにより、エミュレーションのオーバヘッドを最小限に抑えることができ、物理ハードウェア上での動作時と遜色の無い性能を引き出すことが可能となる。また、Xen のマイグレーション機能を利用することによって、ジョブ実行中の PC を途中中断し他の PC に移行させることで、引き続きジョブを行うことができる。もし起動中の PC がすべて利用中でありマイグレーションできない場合でも、休止中の PC を Wake On Lan で起動することで新たにマイグレーション先を確保することができる。

3. 暗号化によるグリッドエージェント間通信

   PC クラスタ内では、高速にかつ密に通信を行う MPI などが十分に性能を発揮できるように、セキュリティレベルは低く設定しなければならない。しかし、インターネットを介する際には、他者からの介入や安全保持のためにセキュリティレベルは高く維持する必要がある。そのため、本システムでは通信に暗号や認証技術を利用して安全にリモートコンピュータと通信できる SSH(Secure Shell) を採用することによって通信経路の安全を確保する。SSH によって通信経路全体の暗号化（トンネル化）を行うことで、通信内容の漏洩を防ぐ。また、SSH による公開鍵認証を利用することにより、第 3 者による不正侵入を防止する。

   また、各グリッドエージェントへの外部からの通信は、SSH が使用するポート以外は遮断し PC クラスタ内の安全を維持する。しかし、これではグリッドエージェント間での通信も遮断されてしまい、クラスタ間で相互に通信することができない。そのために、各クラスタ間で通信が必要となるときには、SSH の機能の一つであるポートフォワーディングを用いてグリッドエージェントが TCP 接続を暗号化し、グリッドエージェントを介してクラスタ間の通信を行う。これにより、必要最小限のポートのみを必要最小限の時間解放することで、インターネット側のセキュリティを高く保持したまま通信することが可能となる。

5 おわりに

インターネット（WAN）上に分散されている PC クラスタをつなぎ、それらを一つのシステムと考えた分散 PC グリッドシステムを提案した。本システムでは、各 PC クラスタに管理サーバとしてグリッドエージェントを配置し、これらのグリッドエージェントが情報交換することにより、効率の良い計算資源配分と通信負荷の軽減を実現させる。本システムは、現在開発中のため性能評価については今後の課題であるが、セキュリティを保証した上で通信負荷や通信遅延の低減が実現できれば、PC グリッドシステムのための効率の良い計算資源配分手法の一つとなると期待している。

謝辞

本研究は文部科学省私立大学学術研究高度化推進事業 (学術フロンティア推進事業) による助成を受けて行った研究成果である。

参考文献

[1] 合田憲人, 関口智嗣編著. グリッド技術入門. コロナ社, 2008.

[2] 榎原博之, 森川浩明. “仮想計算機を用いたPCグリッドの開発”. RCSSディスカッションペーパー　関西大学ソシオネットワーク戦略研究センター, 第79号, Jan. 2009.

[3] Ian Foster. “What is the Grid? - a three point checklist”. *GRID today*, Vol. 1, No. 6, July 2002.

[4] Ian Foster and Carl Kesselman. *The Grid 2*. Morgan Kaufmann, November 2003.

[5] グリッド協議会. “http://www.jpgrid.org/index.html”.

[6] Hyper-V. “http://www.microsoft.com/japan/windowsserver2008/technologies/hyperv.mspx”.

[7] ITpro編. すべてわかる仮想化大全2009. 日経BP, 2008.

[8] ITpro編. すべてわかる仮想化大全2010. 日経BP, 2009.

[9] 森川浩明, 榎原博之, 大西克実, 中野秀男. “仮想計算機を用いたジョブマイグレーションのPCグリッドへの適応”. 情報処理学会 数理モデル化と問題解決（MPS）研究報告, Vol. MPS73, No. 5, pp. 17–20, March 2009.

[10] Parallels. “http://www.parallels.com/”.

[11] SETI@home. “http://setiathome.berkeley.edu/”.

[12] 曽山豊. “企業におけるグリッド・コンピューティングの活用とその成果”. グリッド協議会セッション, Grid World 2006, 2006.

[13] VMware. “http://www.vmware.com/”.

[14] Xen. “http://www.xen.org/”.