

社会科学アプリケーションへの 広域仮想クラスタ構築技術の応用に関する一考察

伊達進・市川昊平・楠元康之



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

社会科学アプリケーションへの 広域仮想クラスタ構築技術の応用に関する一考察

伊達 進* 市川 昊平† 楠元 康之‡

概要

近年、仮想計算機技術を応用した広域分散計算技術、すなわち、クラウドコンピューティング技術への関心と期待が様々な産業および学術分野において急速に高まりつつある。このことは社会科学分野においても例外ではなく、クラウドコンピューティング技術への期待は高い。本研究では、そのような背景から、われわれが新しいクラウド技術を目的として研究開発を進めている、広域環境上の遊休資源を動的に集約し、仮想的な計算クラスタを構築可能とする技術を、実際の社会科学アプリケーションへ応用し、クラウド技術の社会科学分野への応用可能性を模索する。さらに、社会科学ならびに情報工学の両分野から今後解決すべき課題を明らかにする。

キーワード：広域仮想クラスタ、クラウド、社会科学アプリケーション

* 関西大学ソシオネットワーク戦略研究センター・研究員
関西大学ソシオネットワーク戦略研究機構・機構研究員
大阪大学サイバーメディアセンター・准教授
大阪大学大学院情報科学研究科・准教授

Email: date@cmc.osaka-u.ac.jp

† 関西大学ソシオネットワーク戦略研究機構・機構研究員
関西大学ソシオネットワーク戦略研究センター・研究員
大阪大学情報基盤推進本部・助教

Email: ichikawa@cmc.osaka-u.ac.jp

‡ 大阪大学大学院情報科学研究科・博士前期課程2年

Email: kusumoto.yasuyuki@ais.cmc.osaka-u.ac.jp

A View of Cloud Computing Technology in Social Scientific Application

Susumu DATE[§] Kohei ICHIKAWA^{**} Yasuyuki KUSUMOTO^{††}

Abstract

Recently, wide-area distributed computing technology leveraging virtual machine technology, that is, cloud computing technology increasingly and rapidly has been gathering attentions and expectations from a variety of industrial and academic fields. This trend is not exceptional in social scientific field. For the background described above, we have been working on the development of a new cloud technology that allows scientists to aggregate idling computational resources on the Internet and then build a user-dedicated virtual cluster using virtual machine technology. In this paper, we explore the feasibility and potential of cloud computing technology in the social scientific field, by actually applying our cloud solution to actual social scientific applications. Furthermore, clarification of future issues to solve from both social scientific and information technological perspectives is aimed in the paper.

Keywords : Wide-area virtual cluster, Cloud computing, Social scientific application

[§] Research Fellow, The Research Center of Socionetwork Strategies, Kansai University
Researcher, The Research Institute for Socionetwork Strategies, Kansai University
Associate Professor, Cybermedia Center, Osaka University
Associate Professor, Graduate School of Information Science and Technology, Osaka University
Email: date@cmc.osaka-u.ac.jp

^{**} Research Fellow, The Research Center of Socionetwork Strategies, Kansai University
Researcher, The Research Institute for Socionetwork Strategies, Kansai University
Assistant Professor, Office for Information Infrastructure Design, Osaka University
Email: ichikawa@cmc.osaka-u.ac.jp

^{††} Graduate Student, Graduate School of Information Science and Technology, Osaka University
Email: kusumoto.yasuyuki@ais.cmc.osaka-u.ac.jp

1 はじめに

近年、急速にプロセッサ性能が向上しつつあり、物理計算機上で仮想的な計算機を稼働させる仮想計算機 (Virtual Machine) 技術が現実的に利用可能となりつつある。実際、今日においては、VMWare、VirtualPC 等の商用ソフトウェアを用いて、仮想計算機技術についての詳細な知識や技術を有さないユーザが、Windows や Mac OS が稼働する計算機上に、Windows や Linux などのオペレーティングシステムで稼働する仮想的な計算機を複数台容易に稼働させることが可能である。また、Xen [1] や KVM (Kernel-based Virtual Machine) [2] 等の様々な仮想計算機技術がオープンソースソフトウェアとして開発が進められている。

一方、今日までのネットワーク技術の発展は、分散、並列、協調計算技術の発展へとつながっている。同一組織上で管理運用する複数の計算機を連携駆動させ、分散並列計算を可能にするクラスタコンピューティング、さらに広域環境上で複数の組織に分散する、クラスタやスーパーコンピュータを動的に統合し、大規模計算を可能にするグリッドコンピューティングは、そのような分散、並列、協調計算技術の代表例である。現在では、わが国においても、SINET3 [3] や JGN2plus [4] などの 10Gbps 級の広帯域学術ネットワークがすでに利用可能であり、これらを利用した広域分散計算に関する研究開発が活発化している。

このようなプロセッサ性能向上による仮想計算機技術の発展や広域分散計算技術の発展は、それら技術の融合を通じて、新しいコンピューティング技術、すなわちクラウドコンピューティングにむけた技術およびその応用に関する研究開発潮流を引き起こしている。事実、今日においては、Amazon 社の提供する Amazon Elastic Compute Cloud (Amazon EC2) [5]、米国カリフォルニア大学サンタバーバラ校らの研究グループが中心となって研究開発を推進している Eucalyptus [6]、シカゴ大学およびフロリダ大学の研究グループを中心に研究開発が進められている Nimbus [7] 等のクラウド技術をはじめとして、仮想計算機技術と広域分散計算技術の融合による新しいコンピューティングスタイルと技術の確立を目的とした、クラウド環境構築技術およびその応用に関する研究開発が世界中で推進されている。さらにいえば、そのような研究開発潮流は、クラウド環境とクラウド環境を接続するスカイコンピューティング [8]、計算を自組織で所有する計算資源で行い突発的に発生する計算ニーズに対応するため必要に応じてクラウド環境上の計算資源を利用しようとするサージコンピューティング、あるいはハイブリッドコンピューティング [10] といった新しいコンピューティングスタイルの模索へとつながりつつある。

一方、社会科学分野では経済や金融、政治、福祉における社会現象を解明する手段として、その社会現象に関わる人間の行動をモデル化し、そのモデルを特徴付けるパラメータを現実の統計データやアンケート調査などによって集計されたデータを用いて推定するという手法が行われている。このような現実に観測されるデータからのパラメータ推定においては、その推定精度向上のために、ランダムな初期値からの多数の推定を行い、その結果が一意に収束するかどうかをみるモンテカルロ法による推定結果の精緻化が一般的に行われる。このような多数の推定を繰り返す手法においては、近年の情報技術の発展に伴う取り扱うデータ量の増加や、用いる推定手法の高度化による計算時間の増大が大きな問題となっている。そのため、計算時間の短縮を目的とし、多数の繰り返し推定計算を複数の計算機を同時に並列に用いることによって短時間で計算可能とする分散並列計算に対する期待が社会科学の分野においても高まっている。

本研究では、上述した情報工学分野における研究動向背景に着目し、クラウドコンピューティング技術の社会科学分野への応用可能性を模索するとともに、社会科学ならびに情報工学の両分野から今後解決すべき課題を明らかにすることを目的とする。特に、本稿では、われわれが研究開発を進めているクラウド環境構築のための仮想クラスタ技術を用いて構築した仮想クラスタ上で、実際の社会科学アプリケーションを実装し、その計算性能、データの安全性の観点から実用性および有用性を評価する。

本論文の構成は以下の通りである。2節ではまずわれわれが研究開発を進めている仮想クラスタ構築技術について概説する。3節では、本稿が取り扱う社会科学アプリケーションについて説明する。4節では、2節で説明する仮想クラスタ構築技術を用いて構築した仮想クラスタ上で、3節で紹介した社会科学アプリケーションを実装し、その計算性能、データ安全性について評価を行う。5節で今後の課題について整理し、6節で本稿をまとめる。

2 仮想クラスタ構築技術

本節では、われわれが研究開発する仮想クラスタ構築技術について概説する。以下では、まずわれわれが研究開発を進める仮想クラスタ構築技術の開発背景を説明し、その後仮想クラスタ構築技術について詳細に説明する。

2.1 開発背景

1 節で記したように、今日では、仮想計算機技術と広域分散計算技術の融合により、Nimbus、Eucalyptus などのクラウド構築技術が利用可能となりつつある。しかし、これらのクラウド構築技術の多くは、単一組織内の仮想化された計算資源を動的に集約し、ユーザに仮想クラスタ資源を提供するものの、異なる管理ポリシーを有する複数の組織に分散する仮想化された計算資源を動的に集約した仮想クラスタの構築には課題が残されている。

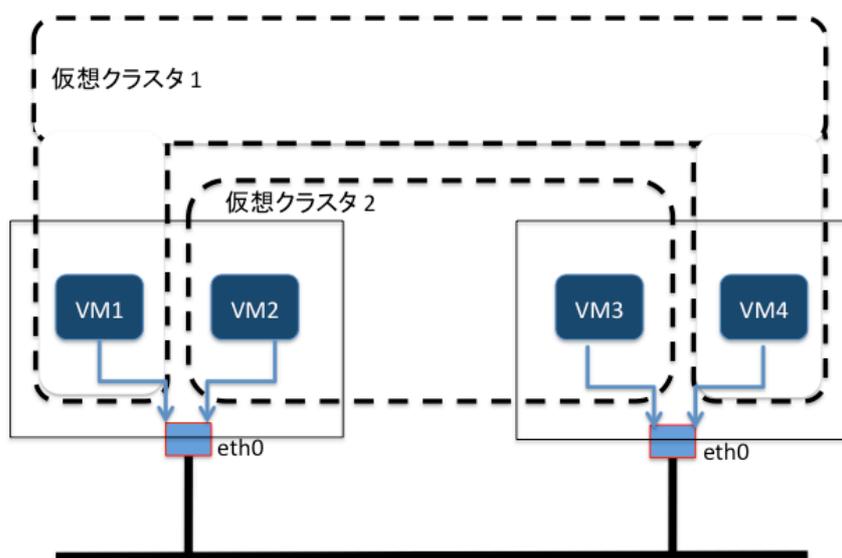


図 1: 仮想計算機間でのネットワークインタフェース共有

図 1 に物理計算機上で 2 台の仮想計算機 (VM) が稼働している状況を示す。一般的に、物理計算機上で稼働する仮想計算機間を接続する場合、図 1 が示すように、同一物理計算機上で稼働する仮想計算機は、仮想計算機が稼働する物理計算機のネットワークインタフェースを共有する。例えば、図 1 では、仮想クラスタ 1 が VM1 と VM4、仮想クラスタが VM2 と VM3 から構成されているが、クラウド環境ではこれら仮想クラスタはそれぞれ異なるユーザによって利用されることが一般的である。そのため、仮想クラスタ 1 のユーザが、VM 1 を通じて物理計算機 A のネットワークインタフェースを流れるトラフィックを監視することで、仮想クラスタ 2 の通信内容を盗聴することも可能となる。そのため、異なる物理計算機上で稼働する仮想計算機を集約し仮想的なクラスタを構築する場合、それぞれの仮

想クラスタが利用するネットワークをセキュリティ上の観点から分離する必要がある。

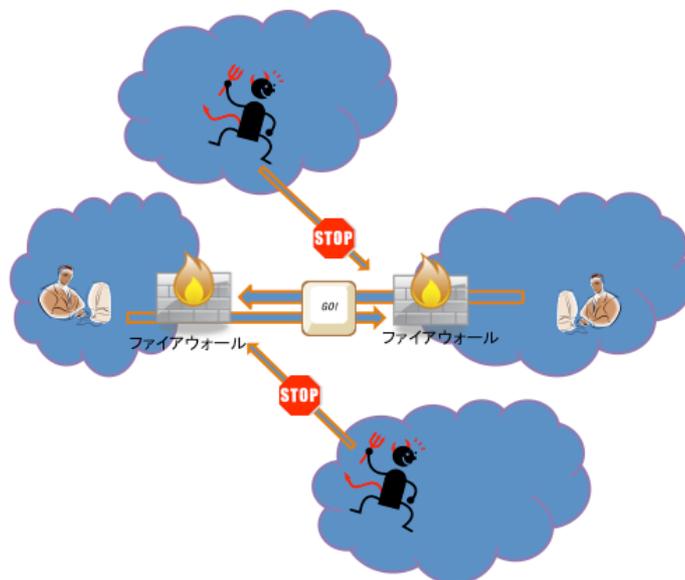


図 2: ファイアウォールによるネットワークセキュリティ

このような観点から、Eucalyptus などの多くのクラウド技術では、静的あるいは動的な方法で仮想クラスタごとに VLAN (Virtual Local Area Network) を形成することで、ネットワークの分離を行い、仮想クラスタ間のセキュリティを確保する技法を採用している。しかし、VLAN は同一管理ポリシーを有する単一組織内においては静的な方法で形成できるが、異なる管理ポリシーを有する複数の組織間においては困難である場合がある。例えば、図 2 に示すように、今日においては組織内のネットワークへのトラフィックをフィルタリングするファイアウォールを設けることが一般的である。この場合、複数の組織に分散する仮想計算機を集約する仮想クラスタを構築するためには、あらかじめファイアウォールの設定を当該組織の管理者間で行っておく必要がある。

このような考察から、われわれは、物理ネットワークの構造によらず論理的なネットワークを形成するオーバーレイネットワーク技術に着目し、広域環境上に分散する仮想計算資源をオーバーレイネットワーク上で動的に集約し、ユーザ占有の仮想クラスタを構築する技術の研究開発を推進している。より技術的には、われわれが研究開発を進める仮想クラスタ構築技術は、物理計算機上で仮想計算機技術 Xen で実現される仮想計算機間を、オーバーレイネットワーク技術 PIAX (P2P Interactive Agent eXtensions) [10] によって実現するオーバーレイネットワーク上で接続することにより、仮想クラスタを提供する。

2.2 コンセプト&アーキテクチャ

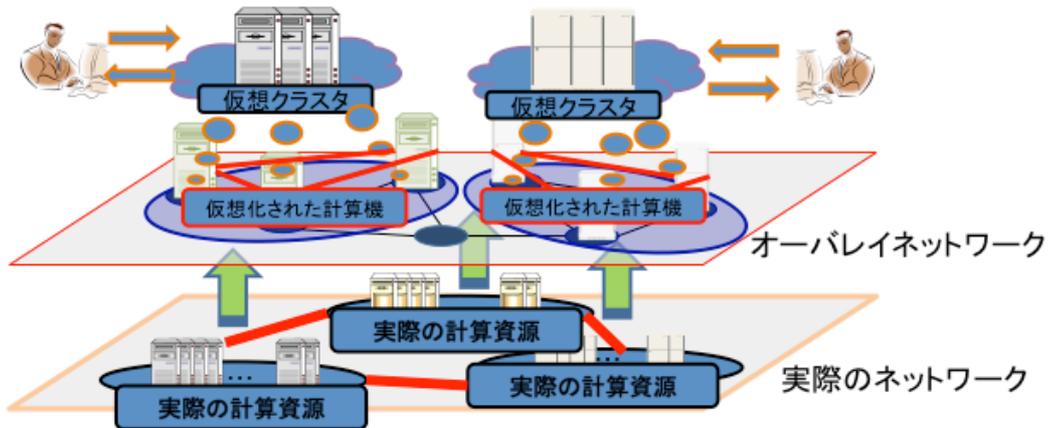


図 3： 仮想クラスタ構築技術

図 3 にわれわれが開発を進めている仮想クラスタ構築技術のコンセプトを示す。図 3 に示すように、われわれが研究開発を進める仮想クラスタは、広域に分散する物理的な計算機をオーバーレイネットワーク上で動的に集約し、仮想計算機技術を応用することで、仮想的に単一クラスタシステムを構築する。このように仮想クラスタを構成することで、ユーザは広域に分散した計算機をあたかも通常のクラスタを操作するかのように使用することができ、利便性の高い広域分散計算環境をえることができる。以下では、われわれが開発を進めている仮想クラスタ構築技術について説明する。

本研究では、仮想計算機技術として Xen を応用する。Xen は複数の仮想計算機をある単一の物理計算機上で動作可能にする仮想計算機技術の一つである。仮想計算機は物理計算機上に物理計算機（ホスト計算機）とは独立して構築可能な仮想的な計算機（ゲスト計算機）であり、ユーザは仮想計算機上に独自にオペレーティングシステムやライブラリ等を導入することができる。

一方、仮想計算機間を接続するオーバーレイネットワークの確立のために、本研究では、大阪大学で研究開発が進められている PIAX を採用する。PIAX は、物理ネットワーク上に物理ネットワークとは独立して、組織間にまたがる仮想的なネットワーク（オーバーレイネットワーク）を形成し、メッセージベースの通信を容易に実現できる。本研究では、このメッセージベースの相互通信を応用し、広域に分散する仮想計算機間を接続するプライベート

なオーバーレイネットワークを構築することで、仮想クラスタを利用するユーザにとって安全なネットワークを提供することをねらっている。

本研究でわれわれが研究開発を行っている仮想クラスタ構築技術では、PIAX が形成するオーバーレイネットワーク上で、仮想計算機間で発生する IP (Internet Protocol) 通信を、PIAX の提供するメッセージベースの通信機構により中継することで、仮想計算機間の通信を確立する。より具体的には、PIAX では PIAX が確立するオーバーレイネットワーク上に一意に特定可能な識別 ID をもつエージェントが配置され、この識別 IP を用いたメッセージベースの通信が行われるが、本研究で開発した IP 通信中継機構 [11] では仮想計算機から発生する IP パケットを、当該仮想計算機上に配置された PIAX エージェントで PIAX のメッセージフレームにカプセル化し、目的の仮想計算機に配置されたエージェントに配送することで、仮想計算機間の IP 通信を実現する。

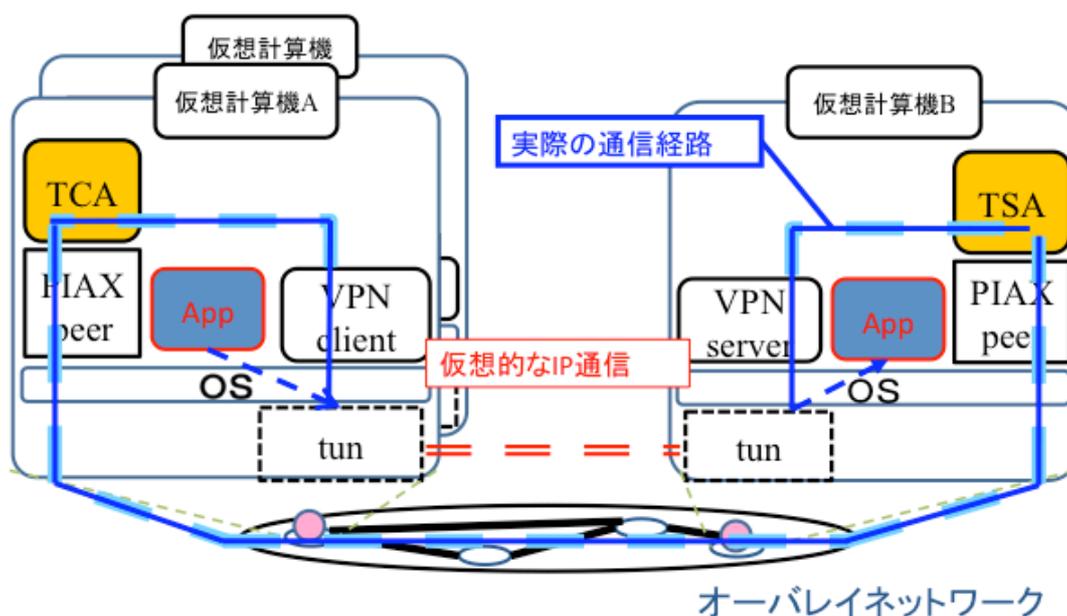


図 4: 通信中継機構

さらに、IP 通信中継機構では、図に示すように、PIAX の提供する API (Application Programming Interface) を用いて、TCA (Tunnel Client Agent) および TSA (Tunnel Server Agent) を開発し VPN 技術と連携させることにより、仮想計算機間で安全なプライベートネットワークの実現を行う。図 4 に通信中継機構のアーキテクチャおよび動作を示す。このとき、仮想計算機 A, B にはそれぞれ VPN クライアント、VPN サーバがインストール

ールされており、A、BともPIAXが提供するオーバーレイネットワークに参加していることを前提とする。本通信中継機構では、(1) 仮想計算機BにおいてTSAはVPNサーバにくるコネクション確立要求を待ち受ける一方、(2) 仮想計算機AではVPNクライアントがVPNの接続をTCAに要求する。この後、(3) 接続要求を受け取った仮想計算機AのTCAは仮想計算機BのTSAにPIAXオーバーレイネットワークを介して通信を行い、接続要求を配送する。その後、(4) TSAが接続要求をVPNサーバに転送し、(5) VPNサーバはその接続要求を受け取る。このような動作により、VPNクライアントからVPNサーバに接続要求が送信され、仮想計算機A、B間でVPNが確立される。その後、一旦VPN接続が確立されると各計算機には仮想ネットワークインタフェース(tun)が構成され、当該インタフェースを通じて仮想計算機間で相互にIP通信が可能となる。

3 アプリケーション

2節では、われわれが研究開発を推進している、仮想クラスタ構築技術、特にその中核となるオーバーレイネットワーク技術を応用するIP通信中継機構について記した。本論文1節で上述した通り、本論文では、クラウドコンピューティング技術の社会科学分野への応用可能性を模索するとともに、社会科学ならびに情報工学の両分野から今後解決すべき課題を明らかにすることを目的とする。本節では、本目的のために、本研究で利用する社会科学アプリケーションについて概説する。

3.1 アプリケーションの概要と計算特性

本研究で用いる社会科学アプリケーションは観測可能な統計情報からメタヒューリスティクスな手法を用いて、直接観測不可能な社会モデルを推定するプログラムである。計量経済学においては、未知の社会モデルを観測された統計情報から統計的に推定することは一般的な問題である。推定手法には様々な方法が存在するが、その手法が確率に依存する手法であるかどうか注目すると、大きく分けて2つに分類される。利用できる観測データが十分にある場合やモデルが比較的単純な場合は多重回帰やロジスティック回帰分析が用いられ、推定値は観測データから一意に決定される。一方で、観測データが少ない場合は、ベイズ推定手法の一種であるマルコフ連鎖モンテカルロ法が用いられ、モデルが複雑すぎて一般的な解法が見つからない場合は、メタヒューリスティクスな手法が用いられ、推定値は確率に応じて一意に定まることはない。後者の手法の場合、推定結果の精度向上のため、さらにモンテカルロ法による精緻化を行うことが一般的であり、膨大な計算量を

必要とする。本研究では、社会科学アプリケーションにおける社会モデル推定の必要性和、計算量を必要とし大規模計算環境を必要とするという視点から、後者の手法を用いるアプリケーションを選択した。

本研究で実際に扱うアプリケーションは噂などの情報が人から人へと広がる際の基盤となる社会ネットワークを推定するプログラム **SN_Rebuilder** [12] である。社会ネットワークの推定は金融不安に関する情報伝播や年金未納行動の伝播など、人々の間で交換される情報が社会現象を形成するようなモデルにおいて非常に重要な問題である。社会ネットワークは直接観測することは不可能であり、間接的に観測可能な情報から推測することしかできず、観測される情報から解析的に推定できるような問題でもない。**SN_Rebuilder** はアンケート調査などにより観測可能な度数分布やクラスタリング係数、媒介性などから、メタヒューリスティクスな手法の一つである遺伝的アルゴリズム (GA) を用いて、基となる社会ネットワークの推定を試みるプログラムである (図 5)。現在の実装は、度数分布のみを用いて推定する。

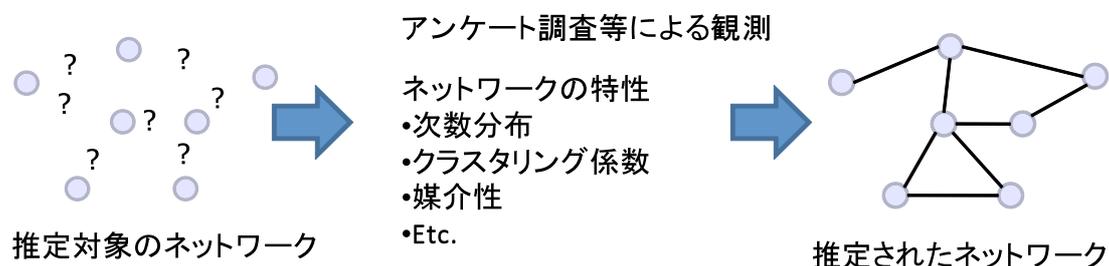


図 5: 社会ネットワークの推定

GA によって導かれる解は使用する乱数に応じて常に変化するため、乱数生成に使用するシードを変化させてモンテカルロ法による結果の精緻化を行う必要がある。GA の計算は、モデルが複雑になるに従って、解の収束までに時間が掛かるが、それを複数回実行しなければならないとなると、さらに多大な計算量を要求することになる。ただし、このモンテカルロ法の過程は全く独立に並行して実行することができるため、複数の計算機において分散計算することによる高速化が期待できる点である。社会科学アプリケーションにおいては、このように複数の計算機間での通信量が少なく、複数の計算機で分散して計算して高速化が期待できるものが多い。

しかし、一方でこのような社会科学が扱うデータはアンケート調査結果など重要な個人情報

報を多数含んでおり、セキュリティの観点から、ネットワークを介してこのようなデータを共有し、複数の計算機で分散計算を行うことに対して大きな懸念が生じている。このような背景から、セキュリティを確保しつつ、取り扱う社会科学の問題に応じて適切な量の計算機資源を確保し、分散並列計算環境を構築する技術に対して、大きな期待が寄せられている。

4 評価

本節は、われわれが研究開発を推進している仮想クラスタ技術を用いて LAN 環境上に構築した仮想クラスタ上で、3 節で記した社会科学アプリケーション SN_Rebuilder を実装し、その計算性能について評価する。その後、データの安全性について考察する。

4.1 性能評価

社会科学アプリケーション SN_Rebuilder の計算性能評価を行う。性能評価環境としては、今回は同一 LAN 上に接続された計算機 7 台を用いる。清算性能の評価に至っては、7 台の計算機をクラスタ（物理クラスタ）として構成した時、および、それら 7 台の計算機上にわれわれの仮想クラスタ構築技術を用いて構築した仮想クラスタとして構成した時の SN_Rebuilder の計算性能を評価する。物理クラスタならびに仮想クラスタを構成する各計算機の性能については、下記の表 1 の通りである。なお、7 台の物理あるいは仮想計算機をクラスタとして構成する際、いずれも 1 台をマスタ、6 台をワーカとし、マスタからワーカへと SN_Rebuilder 内のモンテカルロ法の計算負荷を分散する構成としている。なお、スケジューラには、今回 Torque を採用した。

表 1: 性能評価に用いた物理クラスタおよび仮想クラスタ構成計算機

	物理計算機	仮想計算機
OS	Cent OS 5.3	Cent OS 5.3
プロセッサ	Xeon 3.20 GHz	Xeon 3.20 GHz
メモリ	3 GB	1 GB
ネットワーク	1000BASE-T	1000BASE-T

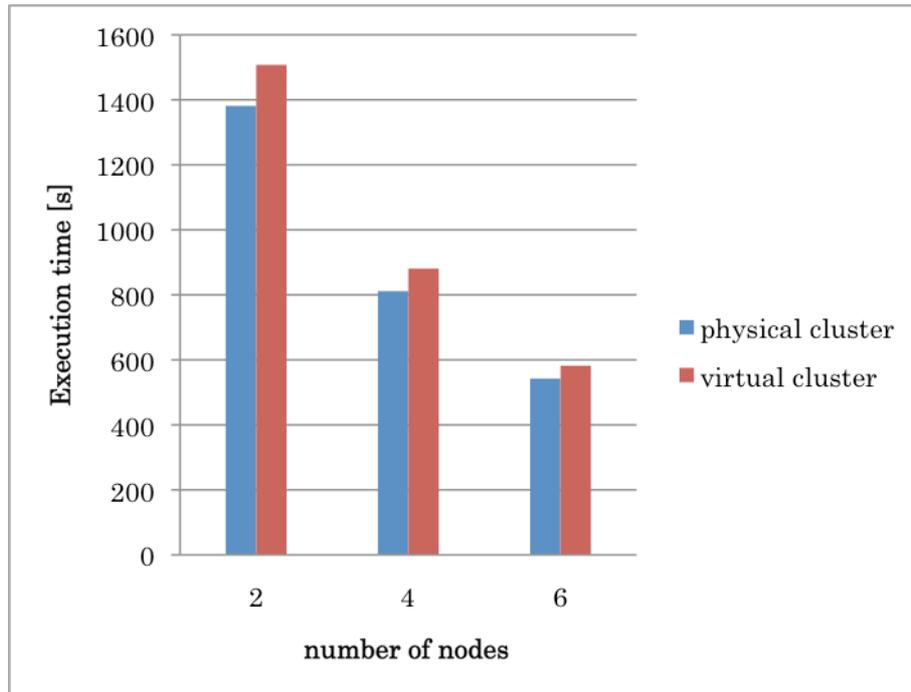


図 6: 物理クラスタおよび仮想クラスタ上での SN_Rebuilder 実行時間

図 6 は SN_Rebuilder を物理クラスタおよび仮想クラスタで実行した時の実行時間をグラフに示したものである。計測実験においては、クラスタを構成する計算機数を、物理クラスタおよび仮想クラスタいずれの場合においても 2、4、6 と変更して実行時間の計測を行った。このグラフから、いずれの場合においても、物理クラスタの方が仮想クラスタと比較して高い計算性能を有していることがわかる。しかし、その一方、仮想クラスタの性能低下は、物理クラスタと比較して約 9%以内に収まっていることもまた示されている。

4.2 データの安全性

4.1 節の性能評価実験を通じて、本論文で提案する仮想クラスタ構築技術によって構築した仮想クラスタは、物理クラスタと比較して約 9%程度の性能低下があることが示された。このことは主として仮想計算機技術 Xen を用いること、仮想計算機間を接続する技術としてオーバーレイネットワーク技術 PIAX を用いること、および仮想クラスタ構成計算機間でのプライベートネットワーク構築のために OpenVPN を用いることに起因するオーバーヘッドから説明できる。しかし、これらの技術は同時にデータセキュリティを確保する上でも有用であり、本研究で実現する仮想クラスタの実用性および有用性を評価する上で計算性能低下とデータ安全性の確保の間にトレードオフがある。

仮想計算機技術 Xen を用いることにより、物理計算機とその上で稼働する仮想計算機のファイルシステムをそれぞれ独立に運用することができる。そのため、われわれが研究開発を進めている仮想クラスタ構築技術を用いてユーザ占有型の仮想クラスタを構築した際には、その仮想クラスタ上で実行する計算で利用するデータを、物理クラスタを利用するユーザからアクセスされることを防ぐことができる。また、オーバレイネットワーク技術 PIAX を用いてユーザ単位でオーバレイネットワークを確立し、かつそのオーバレイネットワーク上で、SSL (Secure Socket Layer) ベースの通信暗号化を行う VPN 技術を用いることで、仮想クラスタを構成する仮想計算機間での通信の秘匿性を確保することができる。そのため、本稿で示した仮想クラスタ構築システムで実現する仮想クラスタ上で実施する計算の安全性は比較的高いといえる。

しかし、仮想クラスタ上で計算を実施するには、計算で利用するデータを実際に仮想計算機上に配置 (ステージイン) し、計算終了後に計算結果を仮想計算機から取得 (ステージアウト) しなければならない。また、個人のプライバシーに関連するような秘匿性の高いデータを利用する社会科学アプリケーション計算においては、その計算終了後に、データおよび計算結果を仮想クラスタから確実に削除し、また仮想計算機そのものについても削除するといった対策も必要になると考えられる。本研究で提案した仮想クラスタ構築システムでは、このような仮想クラスタへのデータステージイン、仮想クラスタからのデータステージアウト、および計算終了後の処理等の運用面については対応ができていない。今後仮想クラスタ技術を社会科学での応用を進めていく際には、このような運用面での課題を十分に検討していく必要がある。

5 今後の課題

4 節で示した性能評価実験より、本研究で用いたような社会科学アプリケーション、すなわちパラメータ分散による分散計算では、仮想化された計算機を複数台用いる場合と、実際の計算機を複数台用いた場合では、その計算性能低下はわずか 9% 以内であることが示された。このことにより、社会科学分野におけるこのようなパラメータ分散型アプリケーションに対しては、仮想計算機技術と広域分散計算技術の融合であるクラウドコンピューティング技術は実用上有効であることが示唆される。

しかし、社会科学アプリケーションにおいても、単純に計算に使用するパラメータで分散できるものばかりではなく、分散された計算間で高頻度かつ大容量にデータをやり取りす

る必要のあるデータインテンシブ型のアプリケーションも少なからず存在する。そのような場合においては、本研究が提案する手法はオーバーヘッドが大きくなりすぎるため、適用は非現実的かつ不可能である。したがって、広域ではなく単一組織あるいは局所的な計算資源のみを用いて、選択的にクラウド環境を構成する、あるいは広域環境下においても広帯域ネットワークに接続された仮想計算機を選択するといったサーバセクション技術が課題となる。

また、データの安全性については、多くの課題が残されている。社会科学分野においては、個人の年収、学歴などのプライバシーに関する機密性が高いデータを扱うことも多い。そのような際に、今日のクラウド技術においては、VLAN や VPN などの技術により、ネットワーク上を流れるデータについては暗号化により第三者からの盗聴から保護することは可能になりつつある。また、広域環境に分散する計算資源を集約する広域分散型のクラウドの構築においても、われわれが研究開発を進める仮想クラスタ構築技術などのように、オーバーレイネットワーク技術を用いる事により、ネットワーク上での保護は容易にできる。しかし、このような機密性の高いデータをクラウド環境へアップロードし、仮想計算機上に配置することは、セキュリティ上大きな課題がある。クラウド環境を構成する仮想計算機はユーザ占有での利用が前提ではあるが、計算に利用されるデータはクラウド環境を提供するプロバイダの管理するデータストレージに他のユーザと一括して管理される可能性がある。この場合、データ自身が暗号化されているか、あるいはそのままであるか、についてもプロバイダの管理方針に依存し、データの機密性の保護は保証されていない。さらに、データ自身がそのようなクラウド環境に一時的にはあれ配置されることが、社会科学アプリケーションで利用するデータ収集時の規約に違反する可能性もある。そのため、医療分野におけるアプリケーションと同様に、匿名化などの手法により、漏洩した場合においても個人を特定できないよう計算前に処置をするなどの徹底したデータ管理は技術面だけでなくユーザの教育といった点でも課題があると考えられる。

6 まとめ

本研究では、近年のクラウドコンピューティング技術の社会科学分野への応用可能性という観点から、われわれが研究開発を進める仮想クラスタ構築技術を実際の社会科学アプリケーションに適用し、その計算性能、データの安全性の観点から実用性および有用性について評価した。その結果、社会科学アプリケーションにも多々みられるパラメータスタディ型の計算問題、すなわち、分散する仮想計算機間で行われる計算プロセスの間に通信が

高頻度に発生しない分散計算問題に対しては、物理クラスタと比較しても仮想クラスタ上でも十分実用的な計算性能を示すことが判明した。また、データの安全性という観点からは、われわれの研究開発する仮想クラスタ技術においては仮想クラスタを構成する仮想計算機間のデータをVPN技術によって暗号化するため、通信の秘匿性は保護されることを本稿で示した。しかし、クラウドコンピューティング環境は永続的に利用するのではなく、一時的にするといった利用方法が主であることから、構築される仮想クラスタ上にデータを配備（ステージイン）、また計算終了後に計算結果をユーザの管理するデータストレージへの移動（ステージアウト）の際、秘匿な通信手法を利用しなければデータの安全性は確保できないこともまた示した。さらに、社会科学アプリケーションへの入力データ、および社会科学アプリケーションの出力データを、どのようにデータストレージで管理し、また同様の目的をもつ研究者間で管理するのかという課題があることもまた本稿で示した。

今後、クラウドコンピューティングは、一つの計算形態として期待と関心がますます高まっていくと考えている。その結果、社会科学分野においても、クラウドコンピューティング技術によってもたらされる計算容量は非常に魅力的であり、有用であると考えられる。しかし、グリッドコンピューティングと同様に、ユーザとなる研究者が管理権限を有していない計算環境に、プライバシーに関わる機密性、秘匿性の高いデータを安全に配備し、計算後に回収することのできる技術の研究開発は、社会科学分野でのクラウドコンピューティング技術の利活用にとって急務であるといえる。

References

- [1] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, Andrew Warfield, “Xen and the Art of Virtualization”, *Proceedings of the nineteenth ACM symposium on Operating systems principles 2003*, pp. 164-177, 2003.
- [2] Kernel Based Virtual Machine (KVM), http://www.linux-kvm.org/page/Main_Page.
- [3] Science Information NETwork 3 (SINET3), <http://www.sinet.ad.jp>.
- [4] JGN2 plus Advanced Testbed Network for R&D, <http://www.jgn.nict.go.jp>.
- [5] Amazon Elastic Compute Cloud (Amazon EC2), <http://aws.amazon.com/ec2/>.
- [6] Daniel Nurmi, Rich Wolski, Chris Grzegorzcyk, Graziano Obertelli, Sunil Soman, Lamia Youseff, Dmitrii Zagorodnov, *Proceedings of 9th IEEE International Symposium on Cluster Computing and the Grid*, pp. 124-131, 2009.
- [7] Nimbus, <http://www.nimbusproject.org/>.
- [8] Katarzyna Keahey, Mauricio Tsugawa, Andrea Matsunaga, Jose A. B. Fortes, “Sky

- Computing”, *IEEE Internet Computing*, vol. 13. no. 5, September/October, 2009.
- [9] Above the Clouds – A Berkeley View of Cloud Computing, <http://berkeleyclouds.blogspot.com/>.
- [10] PIAX (P2P Interactive Agent eXtensions), <http://www.piax.org/>.
- [11] Yasuyuki Kusumoto, Susumu Date, Kohei Ichikawa, Shingo Takeda, Shinji Shimojo, “A communication relay mechanism toward virtual cluster computing on the Grid”, *Proceedings of the 2008 International Symposium on Applications and the Internet*, pp.273-276, 2008.
- [12] Kohei Ichikawa, Toshihiko Takemura, Masatoshi Murakami, Kazunori Minetaki, Taiyo Maeda, “A Novel Approach for Rebuilding Social Network Structure from Web-based Survey: Toward Bank Run Simulation”, *Proceedings of the 8th Annual Hawaii International Conference on Social Sciences*, pp. 833-834, 2009.