

地域インターネット回線仮想常時結合によるメタPCクラスタの構築

—VPNを適用したトンネリング通信によるメタPCクラスタの構築および大規模数値計算実験—

The Construction of a Meta-PC cluster by the Virtual full-time connection
on the Local Backbone Network

—The Construction and the Large-scale Numerical Computation Experiment of a Meta-PC cluster
by the VPN Tunneling Communication Technology on the Local Backbone Network—

森岡 宏行 Hiroyuki MORIOKA

岡山県立大学大学院情報系工学研究科

Graduate School of Computer Science and System Engineering, Okayama Prefectural University

藤野 延竹 Nobutake FUJINO

岡山県立大学大学院情報系工学研究科

Graduate School of Computer Science and System Engineering, Okayama Prefectural University

加藤 隆 Takashi KATO

岡山県立大学情報工学部

Faculty of Computer Science and System Engineering, Okayama Prefectural University

要 旨

大規模な数値シミュレーションを高性能PCで行う場合、CPUの単体性能の限界で処理速度ならびにアクセス可能な実メモリ領域に上限が生じ、取り扱える計算規模に大きな制約を受ける。しかし、多数のPCを高速ネットワークで相互接続してクラスタシステム化することでこの問題は克服できる。さらにシミュレーション対象が複雑構造化して高細精度解析を要すれば、当然、単一のクラスタでも性能限界の壁に当たる。この場合、複数のクラスタをネットワークで結合してメタクラスタ化することができれば、超大規模な数値シミュレーション実行環境が提供できることになる。これは、いわゆるグリッドコンピューティングの基本概念に合致する応用形態の一つでもある。本研究では、この概念に基づき、遠隔地に分散配置されたPCクラスタを、IX中央制御装置を備えた地域ネットワークの下でVPN(Virtual Private Network)接続することで、実用的なメタPCクラスタを構築した。そして有限要素解析法による大規模数値シミュレーション実験を実施し、計算性能に及ぼす地域ネットワークのデータ通信特性の影響を調べ、その有用性の検証を行った。

Abstract

In the case of execution of very large-scale numerical simulations, even if the newest high-grade PC with single CPU is introduced, a performance limit is existed. Then many restrictions of high performance computations will be presented as for the processing speed and the narrow accessible memory area. However, this problem is conquerable by connecting large number of PCs within a high-speed network, and constructing PC cluster systematizations. Furthermore, if a computational problem of engineering simulation encounters complicated structure and requires high-resolution analysis, even a single cluster will also have a performance limitation. In this case, if two or more clusters and network combinations are prepared, thus this system can be carried out and formed into a meta-cluster, then, an executing environment of super-large-scale numerical simulation will build up. This is the basic concept of grid computing. Therefore, the concept was introduced and the practical meta-PC cluster was built by VPN (Virtual Private Network) connection of the PC clusters distributed on three or more remote places through the local backbone network with IX(Internet eXchange) apparatus in this study. Finally, the computational experiment of large-scale numerical simulation based on the finite element method, and evaluation of the data-communication characteristics and verifications of validity of meta-PC cluster were performed.

1. はじめに

最近の情報通信基盤(情報インフラストラクチャー)に関する技術革新はめざましいものがあり、光ファイバ通信路と超高速ルータ機器類の導入によって、広域バックボーンネットワー

ク(以下、WAN: Wide Area Network と略記)のデータ通信速度は格段に向上している。また、いくつかの地方自治体が推進する地域ネットワーク(以下、MAN: Metropolitan Area Network と略記)と呼称される情報幹線網についても飛躍的な進化を遂げている。そして、高速で高品質な情報ネットワー

クを常用できるような環境整備が着実に推進され、高度な利用形態に対応可能^[1]になり始めている。

こうした公的に半無償提供される情報通信基盤の長足の進歩により、従来は高コストな専用回線を敷設しなければ実現が不可能とされてきた、いわゆる広域通信網を背景にした大規模な分散並列処理システム環境、すなわち実用的なグリッドコンピューティングシステム構築を構想する場合、大手の ISP 通信サービス会社 (Internet Service Provider) が提供する一般ユーザ向け汎用サービスの下でも、かなり大規模なシステム開発が低コストに実現できる状況が到来し始めている。

本研究では、広域情報通信網で広範に使用される常用通信手段、すなわち通信装置・通信プロトコル・周辺技術等々の活用を前提に VPN (Virtual Private Network) 技法を導入する。そして、超大規模な数値解析シミュレーション環境の実現のために、MAN の活用を考慮に入れて、地理的に離れた複数の拠点間に設備したコンピュータクラスタ群を一体化した分散並列処理システムの構築を図った。これにより、超大規模な本格的数値シミュレーション実験を実現するための汎用手法、すなわちグリッドコンピューティングシステムのプロトタイプ構築、および広域分散型システム指向の新たな超大規模数値計算諸技法を開発するための基礎研究を行った。

そこで、多数のパーソナルコンピュータ (以下、PC と略記) 群の高速 LAN 結合体、すなわち PC クラスタ (以下、pcCL と略記) について、岡山県内多ヶ所で光ケーブル通信網の枝路が敷設済みの有為な箇所に分散配置した。そして、それら pcCL 群の相互間を、同県が整備している MAN である「岡山情報ハイウェイ」を通して結合し、ベンチスケールのメタ pcCL (以下、MpcCL と略記) の構築と運用化を図った。これによりグリッドコンピューティングシステムを想定した大規模数値計算シミュレーション実験を挙行し、その際の計算性能に及ぼす地域ネットのデータ通信特性の影響を調べ、その評価と実用性の検証を行った。

2. 大規模な並列処理システムについて

膨大な計算量が要求されるという観点から、本研究では、従来はベクトルスーパーコンピュータや共有メモリ型密結合並列コンピュータの利用が一般的であった有限要素法に基づく大規模数値シミュレーションに注目する。こうした分野においては、コスト対性能比の観点から近年重要視されている並列処理システムの最有力手段として、複数の WS (ワークステーション、EWS と同) あるいは高性能 PC 多数を高速 LAN 結合で集結利用するようなクラスタシステムが注目されている。しかし、こうした計算資源でも、今日的な超大規模な計算負荷要求に対しては限界がある。そのため、各方面でクラスタシステムを複合化してより高い性能を引き出すようなメタクラスタ化した分散並列処理システムが採用され始めている。特に汎用的な PC 技術のみを発展的に活用する pcCL 技法は、近年における PC の急速な高性能化と低価格化により、従来のスーパーコンピュータの数十倍から数百倍のコストパフォーマンスを達成し得るよ

うな大きな可能性^[2]を秘めていることが強く認識され始めている。

このため、多数の pcCL を高速広域ネット結合で束ねて本格的なグリッドコンピューティングシステムを構築し、先進的な諸分野を支える基盤技術として広範な大規模計算需要に対し、極小の利用負担経費で高品質な計算実行用の大規模資源を供給できるようにすることが、各方面で重大な関心事となり、その期待は極めて大きい。

2.1 PC クラスタ

図 1 は pcCL の概念図である。市販の汎用 PC 多数を、高速スイッチングハブを介して相互接続し、これら全体を一つの並列コンピュータ資源と見なすもので、各構成要素となる個々のノード PC (以下、nPC と略記) は独自メモリ空間を持つ。こうした並列環境は、分散メモリ式の NUMA システム (Non Uniform Memory Access System) となり、現在主流の IA32 アーキテクチャ (Intel 32bit Architecture) 最大メモリ空間容量内での演算実行という制限^[3]を受けず、理論上ではメモリ空間を無制限に拡大することが可能となる。もちろん、すでに単体 PC 多数を統合化して従来取り扱われたことの無いような大規模問題の求解に挑戦する試みは、メタコンピューティングと呼ばれ、いくつかの研究報告も散見される^[4]。

本研究では、分散配置状態にある複数の pcCL 群を、MAN および大学キャンパス内や特定企業内のイントラネットを利用して相互結合し、それら総体を一組の大規模 pcCL と見なし、こうして拡大化させた大規模クラスタを MpcCL と定義して、その有効活用を図ることとした。

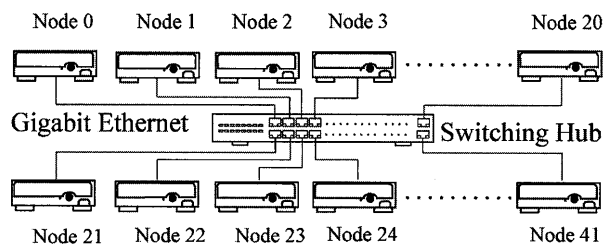


図 1 PC クラスタの概念図

2.2 グリッドコンピューティング

グリッドコンピューティングとは、端的に言えば、これは計算資源の集合化形態に他ならない^[5]。そうした環境下では、エンドユーザは、計算機資源の構成等々を全く気にせずに端末機からジョブを投入すれば、そのネットワーク領域 (ドメイン) 内に接続されている各種コンピュータ群が全体として非常に大きな計算資源になり、依頼した計算結果をより短時間内に受け取ることが可能となる、という基本概念に基づいている。図 2 の概念図に示すように、用語としてのグリッドは、分散配置されているコンピュータ群やデータストレージシステム類、ならびにネットワーク機器類などの諸計算資源を統合化して、あたかも一つの巨大な計算機システムであるかのように再構築することで、当事者間で効率良く運用できるようにするシステム構

成概念であり、将来的には、ユビキタスコンピューティング世界観の中核概念を占める重要技術となるものであるといえる。

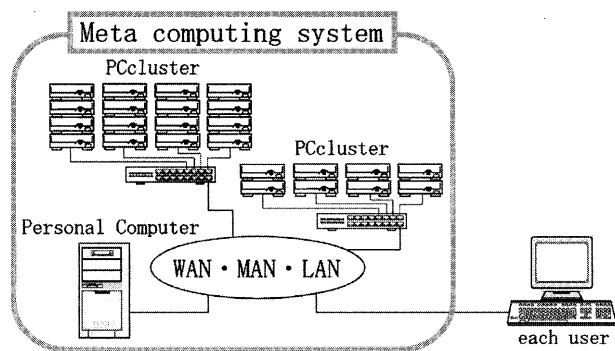


図2 グリッドコンピューティングの概念図

2.3 VPNについて

図3に示すように、インターネット経由で相互に疎な接続関係にある二つのプライベートネットワークの一方と他方との間に、論理空間上で仮想的に整合を図って密結合状態を維持できるようにすることで、全体を一つの閉じたネットワーク領域内と見なせるようにするネットワーク結合技法がVPNである^[6]。もちろん、2者間の関係構築に限らず3者間以上に及ぶ関係構築にも対応できる。

本研究では、常時接続条件下でインターネット上の遠隔地間複数箇所に分散配置されたプライベートネットワーク構成の複数pcCLに対し、このVPN手法を適用することで、全体をMpcCL構成にするようなシステム化を図っている。

インターネット（TCP/IPネットワーク）環境下で任意のnPCを特定する手段としては、32ビット長の既存のIPアドレス（IPv4：Internet Protocol Version 4）、もしくは、128ビット長である次世代のIPアドレス（IPv6：IP Version 6）を利用する方法がある。後者のIPv6は、既存のネットワークIPアドレスが枯渇し始めていることから、今後の急拡大が予想されるIPアドレスの増設要求に対処すべく長期的な要請対応策^[7]として編み出されたものである。しかし、現在のインターネット環境においては、すでに常用されているIPv4が当面は主流の座にあり状況の急変は考えられず、またIPv6が完全に普及浸透し広範に定着するには、かなりの年月期間を要すると思われるため、本報では、IPv4の利用を前提として研究推進を図る。

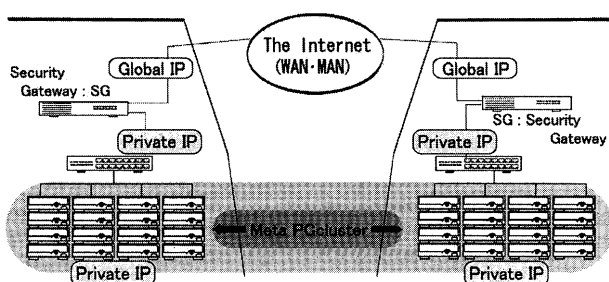


図3 VPNの概念図

2.3.1 IPアドレスについて

インターネット利用のデータ通信の要であるIPアドレスの枯渇対策を鑑みれば、上述したIPv6の新規導入を図る以外の別対策として、既往のプライベートIPアドレス技法^[7]の導入が合理的かつ安定な実用技術であるといえる。

一般に、IPアドレスを使用するホストPC（以下、hoPCと略記）は、①そのhoPCが所属する閉じられたネットワーク領域内でのみ通信を行うものと、②外部ネットワークに対してIPレベルでのアクセスを行うものに大別できる。プライベートIPアドレスは前者①においてのみ使用可能で、言い換えると、プライベートIPアドレスを宛先または発信元とするパケットは、外部ネットワークへの通用的なデータ交信門扉であるゲートウェイ装置を通過することができない。また、後者②のhoPC内ではNIC（Network Information Center）が一元管理するグローバルIPアドレスのみが使用可能である。

2.3.2 IPsec

VPNの実装については様々な手法が考えられ、複数の実用候補が挙げられるが、本研究では、以下のような理由^[8]でIPsecを採用した。

- ・ 技術仕様が標準化され公開されていること、
- ・ VPN拠点毎にSecurity Gateway（IPsec装置：以下、SGと略記）が在れば良く、通信経路上での経路選択スイッチとは既設の汎用ルータで支障が無いため、既往のインターネットをそのまま利用できること、
- ・ 通常のIP通信に対応したセキュリティ機能が提供され、アプリケーションへの変更手続きを要しないこと、

表1は本研究で採用したIPsecの仕様諸元であり、図4は通信データをトンネルモードでIPsec（ESP）化したデータパケットの模式図である。

表1 IPsecの仕様諸元

鍵交換プロトコル	IKE(Internet Key Exchange) /Pre-Shared Key
セキュリティの実装	IPsec(ESP)/ Tunnel mode
認証アルゴリズム	MD5 (Message Digest 5)
暗号化アルゴリズム	DES (Data Encryption Std.)

ここで、トンネルモードによるIPsec（ESP）の通信手順を概説すると、先の2.3.1節で説明した①のネットワークで構成されたpcCLをインターネットに接続してIPレベルでの通信を行う場合、pcCL内の任意のnPCから送信されたプライベートIPアドレスを宛先および発信元とするデータパケット（図4（a）参照）は、

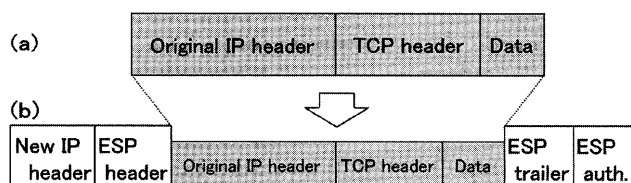


図4 トンネルモードのIPsec（ESP）化パケット

まず SG において下段の (b) のように、カプセル化されてグローバル IP アドレスが記載された新しいヘッダとトレイラが付加されることになる。これに続くインターネット経由のデータ通信過程では、通常のパケットデータと見なされて経路通過し、目的の SG へ配送された後にそこでカプセルが開封されて送信側で付加されたヘッダとトレイラが外される。そして、IP アドレスに合致した目的の nPC へデータ配信されて送信完了となる。こうした一連の IPsec 処理が、トンネルモードによるデータパケット配信、すなわちトンネリングと総称されるインターネット通信技法である。

2.4 地域ネット：岡山情報ハイウェイ

本研究では、岡山情報ハイウェイという MAN 配下において同県下の多ヶ所に分散配置した複数の pcCL を相互接続することで、全体を通信効率が良好でデータ通信の高速維持が可能な VPN 化 MpcCL システムの構築を具体化させ、大規模な数値計算実験でその実証を行う。そのために、この岡山情報ハイウェイの NOC 拠点 (Network Operation Center) において、この情報通信幹線網に直接接続方式の地元 ISP 通信サービス会社の一つである「晴れの国ネット」および「B フレッツ接続サービス (NTT 系光ファイバ通信網)」を採用してインターネット接続を行う。このように、岡山情報ハイウェイを WAN の一部、すなわち Web 上の選択的なバックボーン通信路として利用することで、図 5 に示すように、通常の国内通信トラフィック交換を支配する主に東京都下に設置の国内 IX (Internet eXchange) 経由の多経路多パスステップのデータ交換通信を回避して、地域 IX (Regional IX) 経由の最短経路 (図中の太実線経路) で、極めて効率良い高速データ通信機能⁽¹⁾を確保できる。

図 6 は、広域ネット経由および地域ネット経由の両条件下におけるデータパケット送信時のルーティングパス回数 (リレーステップ数) の実例として、CRT 画面からスクリーンダンプして比較した結果である。この図から分かるように、広域ネット経由のデータ送信に要するリレーパス数に比べ、IX 装置を具備した地域ネット経由のデータ送信に要するリレーパスステップ数が圧倒的に短縮される。本例では、前 3 者平均値の 57.0ms に対して後 3 者平均値が 5.7ms 弱で、ほぼ 10 倍の高速レスポンス性能を示すことが確認でき、大容量かつ高速データ通信に大きく寄与し、グリッドコンピューティングシステムに好適であることは自明である。

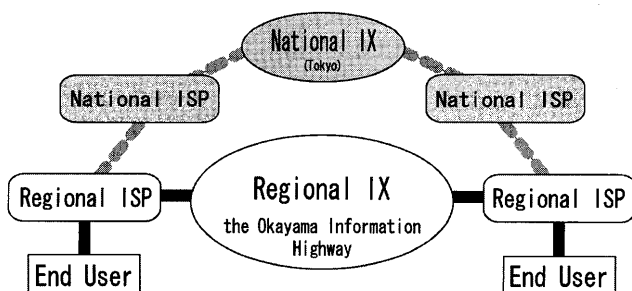


図5 地域IXを経由する短縮パス通信の概念図

```
c:\>tracert 210.167.70.xx

Tracing route to bflets0xx.harenet.ne.jp [210.167.70.xx]
over a maximum of 30 hops:

  1  2ms    1ms    1ms  210.153.207.243
  2  5ms    6ms    3ms  nas932.okayama.nttpc.ne.jp [210.165.247.134]
  3  5ms    6ms    4ms  210.165.247.129
  4  6ms    3ms    6ms  210.153.248.249
  5  34ms   35ms   36ms  210.153.241.81
  6  34ms   36ms   35ms  xg4-0-n-otenachi-core2.sphere.ad.jp [202.239.114.218]
  7  38ms   35ms   35ms  210.173.176.63
  8  *      *      *    Request timed out.
  9  57ms   56ms   54ms  as7664.dix.sakura.ad.jp [61.211.233.28]
 10  56ms   55ms   55ms  fletsntt.harenet.ne.jp [210.167.64.134]
 11  55ms   61ms   57ms  fletsntt2.harenet.ne.jp [210.167.69.3]
 12  58ms   56ms   57ms  bflets0xx.harenet.ne.jp [210.167.70.xx]

Trace complete.

c:\>
```

(a) 広域ネット経由

```
c:\>tracert 210.167.70.xx

Tracing route to bflets0xx.harenet.ne.jp [210.167.70.xx]
over a maximum of 30 hops:

  1  2ms    1ms    1ms  210.167.70.67
  2  4ms    3ms    3ms  fletsntt2.harenet.ne.jp [210.167.69.3]
  3  5ms    6ms    6ms  bflets0xx.harenet.ne.jp [210.167.70.xx]

Trace complete.

c:\>
```

(b) 地域ネット経由

図6 データパケット送信時のルーティングパスの実例

3. 計算実験のシステム環境

3.1 クラスタ計算機の分散配置

本研究では、pcCLを岡山県下の地元企業3社の協力を得て3拠点に分散配置した。各pcCL内は、個々のnPCの中核CPUとして、米Intel社製Celeronプロセッサ(1.7GHz)、半導体高速メモリ1024MBを搭載した同一仕様のPC各々14台ずつを、高速スイッチングハブを介してGigabit Ethernet接続(以下、GethrCN.と表記)かFast Ethernet接続(以下、FethrCN.と表記)の両結合様式を任意に切り換えられるように準備した。各拠点ともにNTT-Bフレッツ接続の光ファイバを敷設導入し、前節(2.3節)で概説したVPN手法を適用し、分散配置した3組のpcCLでMpcCL化を図った。

また一方、単一の研究棟内で上記3組のpcCLをGethrCN.およびFethrCN.で高速LAN結合したMpcCLによる計算実験も進め、両システム形態の差違の影響、ならびに両計算結果に基づく特性間のアナロジー性に関する基礎データを調べ、整理結果に考察を加えた。

3.2 大規模並列計算手法について

任意のnPCにおける計算内容や中途結果が他のnPCの演算過程に頻繁な参照を要するような並列演算を繰り返し行う場合には、必然的に各nPC間におけるデータ通信が高頻度が発生する。こうした計算実行時のプロセス間通信手法としては、Socket通信ライブラリを導入してTCPやUDPプロトコルを直接組込む方法や、MPI(Message Passing Interface)などの並列処理化専用ライブラリ群等を導入することで対処できる。

本研究では、通信APIとして後者の具体的実装である

MPICH^[9]を採用し、膨大なメモリ空間を要する計算対象について検証する。そのために、連続体の塑性変形加工に伴う材料流動数値シミュレーションモデルに適用される剛塑性有限要素法に基づく大規模連立方程式の求解につき、共役勾配法 (CG 法: Conjugate Gradient Method)^[10]を採用した反復計算による多ケースの計算実験を行った。表2は計算手順の大略であり、本手順に基づいて目的の大型連立方程式を解くことで、対象とする連続体の各要素節点毎の変位速度を求める。

以上の手順を進めるためには、大規模な連立方程式 $Ax = b$ の求解を要し、その反復型求解法の一つである共役勾配法のアルゴリズム^[11]は、以下の式 (1)～(7) で表すことができる。

- (1) $p_0 = r_0 = b - Ax_0$
- (2) $\alpha_k = (p_k, r_k) / (p_k, Ap_k)$
- (3) $x_{k+1} = x_k + \alpha_k p_k$
- (4) $r_{k+1} = r_k - \alpha_k Ap_k$
- (5) $\beta_k = - (r_{k+1}, Ap_k) / (p_k, Ap_k)$
- (6) $p_{k+1} = r_{k+1} + \beta_k p_k$
- (7) go to 2

ただし、記号 A , x_k , b , r_k , p_k は各々係数行列、近似解ベクトル、定数項ベクトル、残差ベクトル、修正方向ベクトルを表す。以上の計算式 (1)～(7) を事前設定した収束条件を満たす微小値以下に到達するまで繰り返す。こうした共役勾配法では、上の式 (2)～(6) までの反復回数が全処理時間の大半を支配すると見なせる。なかでも式 (2) 中の行列・ベクトル積計算 Ap_k 部分の計算負荷量^[12]が最大になるため、この部分に最注力した並列処理プログラム開発を行う必要がある。

表2 計算手順の概要

1	入力準備	解析対象を有限個の要素に分割し、各節点と各要素毎に番号をつける。
2	データの入力	節点座標、要素構成、材料の諸物性値、固定条件 (境界条件) 等を与える。
3	要素剛性マトリクスの作成	入力データに基づき各要素毎に要素剛性方程式を作成する。
4	全体剛性マトリクスの組み立て	各要素剛性マトリクスを所定の行と列にインポーズして全体剛性方程式を作成する。
5	境界条件の処理	節点の拘束条件等の境界条件を剛性方程式の中に組み込む。
6	連立一次方程式の計算	全体剛性方程式を解く。

4. 計算実験および実測結果

前章の計算手法に沿って、実動中の MpcCL 環境下で多数組の解析入力データ条件 (以下、入力データモデル、D モデルと略記) を入力して多ケースの数値計算実験を挙行した。特にここでは、前章 (3.2 節) に概説した共役勾配法に基づく数値求解計算を採用したプログラムを、その前節 (3.1 節) に示した

システム仕様の各 pcCL に実装した。ここでは、これら全体を大規模数値シミュレータの実行環境として、本研究の到達目標とする MpcCL システムのベンチスケールモデルと見なし、多ケースの数値計算実験を行っている。

この MpcCL を利用する計算実験では、各 pcCL 間、および各 pcCL 内の各 nPC 間ネットワーク結合性能を選択的に変更し、それらの通信速度性能が全計算処理時間 (以下、 T_{ALL} と略記) に及ぼす影響を詳しく調べる。その条件設定のために、以下では、各 pcCL 内各 nPC 間の結合様式と pcCL 間の結合様式との対比という観点から、両者の様式対比を「結合種 // 結合種」の形式で表すことにした。そこで、FethrCN. → Fast, GethrCN. → Giga のように略記し、WAN 結合に VPN を適用した場合には、Fast//VPN, Giga//VPN, また単一研究棟内で pcCL 間を FethrCN. もしくは GethrCN. にしたシステム設定条件では、各々 Fast//Fast, Giga//Fast, Giga//Giga の3組として、全5組の表記で使い分けた。

4.1 入力データモデル

本報の数値計算実験に用いた入力データモデルは、つぎの表3のような有限要素法に基づく立方体形状の各辺を均等分割して導出したものである。本表で最大規模 $D=62$ では約75万自由度の大規模連立方程式の求解を行うことになる。本報では、この表3の3組のDモデルを入力し、大規模な数値計算実験を個々に複数回ずつ実施し、 T_{ALL} 等の計算結果については、得られた実測データの平均値を採用している。

表3 入力データモデル

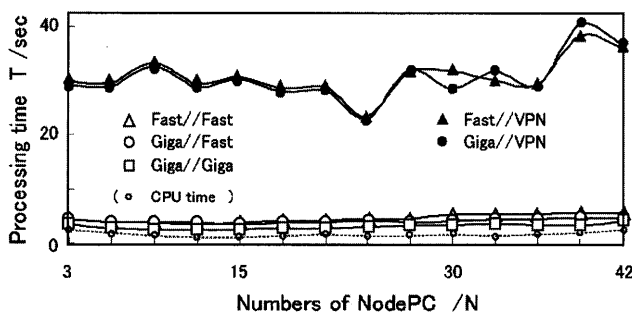
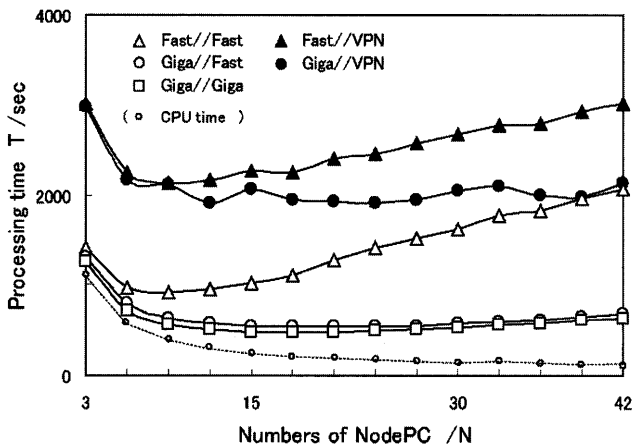
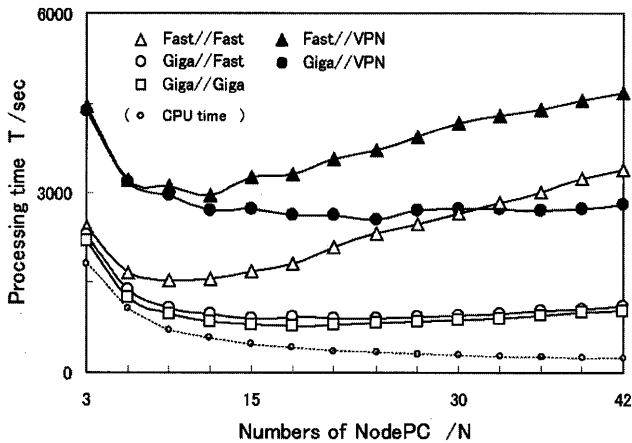
データ名	各辺分割数 (節点数)	総未知数 (自由度)
D=14	14 (15 ³)	10125 (=3×15 ³)
D=54	54 (55 ³)	499125 (=3×55 ³)
D=62	62 (63 ³)	750141 (=3×63 ³)

4.2 総計算所要時間 T_{ALL} の実測結果

ここでは、準備した MpcCL 内外のネットワーク結合様式を前述のように5組に変更し、先の表3に示したデータモデルを入力して数値計算実験を行った。全ての計算実験につき、各 pcCL 毎の nPC 数 N_c が均等になるように $N_c=14$ 、総 nPC 数 (以下、 N と略記) $N=42$ まで逐次増加させ、MpcCL 内の N 数と T_{ALL} 等々との関係を調べた。

これらの計算実験により直接得られる実測値は、 T_{ALL} および各 nPC 毎の CPU 自身が実際に演算処理に占有された正味の総演算所要時間 (以下、 T_{CPU} と略記) であり、各 D モデル毎に分類した整理結果を、それぞれ図7～図9に示した。

これらの結果から、まず $D=14$ の場合は、計算規模が他の2条件よりも圧倒的に小さくて計算処理が数秒から数10秒内の極短時間で終了するため、 $N=3 \sim 42$ の全域において、並列処理の効果やその他の影響が識別しづらく、VPNを採用したWAN結合に伴う通信冗長性の影響が特に顕著に表われてデータ再現性が劣る。

図7 N数と T_{ALL} の関係 ($D=14$)図8 N数と T_{ALL} の関係 ($D=54$)図9 N数と T_{ALL} の関係 ($D=62$)

一方、 $D=54$ の結果は $D=14$ のそれに比べおよそ100倍の T_{ALL} 値になるため、収録データの再現性が安定し分析能も向上する。そして、Fast//VPN、Fast//Fastのシステム設定条件では $N=6 \sim 9$ で、またGiga//VPNでは $N=12$ 前後で最小値になる。

一方、Giga//Fast、Giga//Gigaの場合の T_{ALL} は $N=18$ 前後で最小値になり、 $D=62$ の条件では $D=54$ のそれに比べ約1.8倍を要するが、両者の全体的な特性はほぼ同傾向になるといえる。

こうした結果によると、 T_{ALL} は、全システム内の過半数の通信路がGiga//Fast、Giga//Giga接続のように、GethrCNが主体となるシステム設定では、 T_{ALL} が最小になるN数は、か

なり高N域に移る。またD値がある程度以上の大規模条件であれば、 T_{ALL} の値は、ネットワークの通信性能やシステム設定諸条件その他に依存せず、大略、類似特性になるといえる。

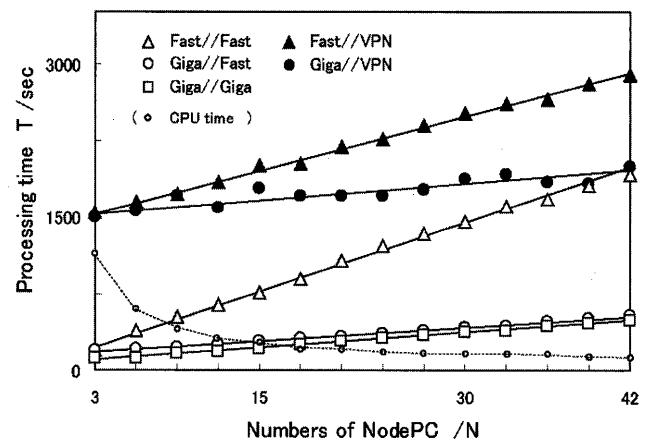
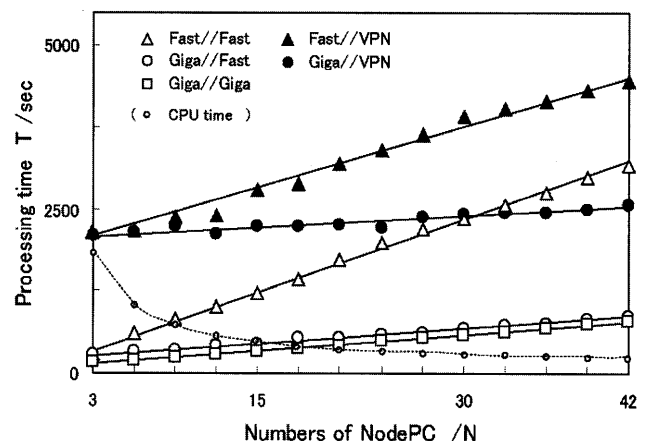
4.3 総計算所要時間 T_{ALL} の構成因子

ここでは、 $D=54$ 、 $D=62$ の両条件下におけるネットワーク通信速度性能の影響を詳しく調べるために、 T_{ALL} から T_{CPU} を差し引いた値がほぼデータ通信処理に要する時間 T_{COM} 、すなわち次式(a)、(b)で近似できるものと見なした。

$$T_{ALL} = T_{CPU} + T_{COM} + T_{etc} \quad (a)$$

$$T_{ALL} \approx T_{CPU} + T_{COM} \quad (b)$$

この T_{COM} を各計算条件下の総通信所要時間と定義し、前出の図8、図9の基礎データを再整理した結果が各々図10、図11である。これらの結果によると、 T_{COM} はN数に比例して単調増加することが明らかで、この T_{COM} 値に律速されて T_{ALL} が決まるような特性と見なせる。また、N数の増加に伴い T_{CPU} が減少し、逆に T_{COM} は増加傾向を示すが、後者 T_{COM} の絶対値が前者 T_{CPU} のそれを上回るようなN数領域になると T_{ALL} が増加に転じる。もちろん、その転換点は T_{ALL} の最小値に一致する。また、入力データモデルの変更に対する影響、すなわちD値への依存

図10 N数と T_{COM} の関係 ($D=54$)図11 N数と T_{COM} の関係 ($D=62$)

性については、N数の全域でD=54とD=62の結果はともに類似特性になるが、高N域における T_{CPU} の低減率はD=54よりもD=62の場合の方が高まる。従って、D値がより大きなデータモデルを計算対象にするほど並列処理効果が強まり、そのまま単調増大傾向が維持されて、 T_{ALL} が最小値を示すN数も、さらに高N域に移ることは容易に推定できる。

5. 計算結果の検討および考察

本報では、MpcCLのネットワーク結合性能を変更した複数のシステム条件下で、反復求解法の一つである共役勾配法の適用を前提にした剛塑性有限要素解析法に基づく大規模な分散並列処理計算実験を実施し、将来的な超大規模な数値シミュレーションに対応可能な基礎特性を把握するための実測データを得た。そして、最大nPC数(=42)を投入した多ケースの計算実験を挙行し、VPN手法を適用したpcCLを3拠点に分散配置させ、それらをMAN接続させてMpcCL化したシステムにおいて、N数と T_{ALL} の関係を詳しく調べることができた。

ここでは、本研究の到達目標である本格的なグリッドコンピューティングシステムによる超大規模数値シミュレーション環境を実現する際のデータ通信特性の予測と超大規模システムとしての実用性を検討するために、WANやnPC間のLANの通信速度性能が T_{COM} 値に及ぼす影響を検討する。そこで、前出の図10、図11に示したN数と T_{COM} との関係に最小自乗法を適用し、次式(c)のような直線近似式で整理し、その結果を表4に示した。

$$f(\chi) = A\chi + B \quad (c)$$

ただし、本研究では大規模領域における計算挙動の分析と特性把握に主眼があるため、いずれもN=12以上の高N領域を対象にした。

この表4の通り、各近似式の勾配Aは、各pcCL内がGethrCN.の場合のMpcCLでは、式①、②および④のような小さな値になり、VPN適用の式④のようにpcCL間の通信性能が低減しても、その影響は僅少である。またD値の増大に対しては、式⑥、⑦および⑨のように1.39倍～1.65倍ほど急傾斜になるが、絶対値としては僅かな変化である。従って、さらに大規模なデータモデルを入力し、より多くのnPC(N数)を投入するような大規模なMpcCLシステムに対しても、充分有意な大規模並列処理化の効果が期待できるものと推定される。

表4 χ (=N数) に対する T_{COM} 特性の直線近似 ($\chi \geq 12$)

接続様式	Func.	D=54	Eq.	D=62	Eq.
Giga/Giga	$f(\chi) =$	$10.4\chi + 60$	①	$17.2\chi + 68$	⑥
Giga/Fast	$f(\chi) =$	$9.2\chi + 143$	②	$14.9\chi + 224$	⑦
Fast/Fast	$f(\chi) =$	$43.5\chi + 138$	③	$72.2\chi + 177$	⑧
Giga/VPN	$f(\chi) =$	$9.8\chi + 1530$	④	$13.6\chi + 1980$	⑨
Fast/VPN	$f(\chi) =$	$34.0\chi + 1460$	⑤	$67.4\chi + 1740$	⑩

また、各近似式の切片値B、すなわちMpcCLの T_{COM} 値については、pcCL内の結合様式が同条件の場合、例えば、式①に対して式②、④、あるいは式③と式⑤とを比較すると、通信性能の低下に伴ってその絶対値の増大が顕著になる。換言すると、通信速度が遅いVPN適用の場合には、pcCL間の通信路が各pcCL内に比べて相対的にかなり低速なため、その低通信路特性に律速されてMpcCL全体の T_{COM} を大きく悪化させ、全体の計算性能を著しく低下させるといえる。

以上の結果のように、N数に対する T_{COM} の直線近似式は、その勾配AがpcCL内の通信路性能を反映し、その切片値Bの相対比がpcCL間の通信路性能の良否を顕著に反映することがわかった。

このような特性については、本報の最大投入N数(=42)よりもさらに大規模な高N数条件に対しても類似結果を得ると推定できる。従って、N数の増加に応じた極大のメモリ空間を確保でき、特大な入力データモデルによる超大規模数値計算条件を想定しても、大容量の実メモリ領域を確保したまま超大規模数値シミュレーション環境の実現を可能にするようなスケールアップ特性を有すると考えられる。

6. おわりに

本研究では、高速ネットワーク接続したPCクラスタを複数拠点に分散配置させ、岡山県が推進する地域ネットワーク「岡山情報ハイウェイ」の潜在機能に着目して、VPN技法を適用したトンネリング通信結合に基づく常時仮想接続式のメタPCクラスタシステムを構築した。そして、計算負荷が重いとされる有限要素解析法に基づく数値シミュレーションに大きなデータモデルを入力して大規模な数値計算実験を挙行することができた。続いて、国内広域バックボーンネットワークの部分構成である地域ネットワークを選択的に活用する本格的なグリッドコンピューティングシステム環境の構築に関する基礎研究を進め、諸特性データを掌握することで、今後の本格システム構築に必要な要素技術の開発とシステム運用等のための基礎知見を得ることができた。

現状では、広域バックボーンネット経由の通信速度は、PCクラスタ内のLAN通信速度に比較してかなり低速であるために総計算処理時間を大きく遅延させる結果になる。もちろん、これを大幅改善する実製品は存在するが、非常に高単価な全ハードウェア回路構成の高級VPNルータ装置の導入を要し、本研究ではこの実機による検証実験に至っていない。しかし、最近の岡山情報ハイウェイについては、先進的な短距離型超高速WDM技術を採用した高多重化双方向通信方式の地域バックボーンネットへの組み替え導入も計画され、その主要情報幹線路の超高速化の再整備も始まっており、1Gbps～10Gbpsの通信速度が常用可能な先進的なMANに転換されつつある。そのため、同一県内というやや近距離内の各所に分散配置された無数のPCクラスタ群に対して、地域ネットという選択的な超高速広域バックボーンネットワークの範囲内で統合化して特色ある本格的なグリッドコンピューティングシステムの構築が

図れ、超大規模数値シミュレーションのための実行環境が実現可能であることが示せたと考えている。

また、国是として進行中の「e-Japan 戦略」、すなわち「全国民が IT 活用の恩恵を享受できる知識創発型の最先端国家社会の実現を目指す大方針」に沿って、国内各地で多くの地方自治体が MAN の整備を積極的に進めている。いずれも岡山県が推進する岡山情報ハイウェイの多線 10Gbps 幹線整備計画や山間過疎地域に対するラストワンマイル光通信路全戸普及計画等の充実したサイバー計画には及ばないが、近隣県である兵庫県、鳥取県、高知県の情報ハイウェイ整備計画⁽²⁾⁽³⁾⁽⁴⁾など、さらに、宮崎情報ハイウェイ 21⁽⁵⁾、福井情報スーパーハイウェイ⁽⁶⁾など、いずれもが 1.8～2.4Gbps という高速光幹線路の構築を前提にした整備計画に基づく本格的な取り組みであると見なせる。そのため、各所で稼働中の幹線路総延長距離の長短にかかわらず、いずれもグリッドコンピューティングシステム環境の構築には非常に好適な潜在的条件を備えていると考えられる。従って、本研究と類似のグリッドプロジェクトの立ち上がり、近隣県の相互連携で超大規模に拡大化させたメタ・グリッドコンピューティングシステム環境に発展させるような将来構想については非常に明るいといえる。

注

- (1) <http://www.pref.okayama.jp/kikaku/joho/joho.htm>
- (2) <http://web.pref.hyogo.jp/ehw/index.htm>
- (3) <http://www.pref.tottori.jp/jouhou/highway/top.htm>
- (4) <http://www.pref.kochi.jp/~jkiban/highway/highway.html>

- (5) <http://www.mjh21.net>
- (6) <http://info.pref.fukui.jp/jouhou/fish/>

参考文献

- [1] 長田智和, 谷口祐治: ブロードバンド通信対応の地域 IX 実現に向けた実証実験と評価, 情報処理学会論文誌 Vol.43 No.11, pp.3509-3518 (2002).
- [2] 藤澤克樹: 大規模最適化問題への挑戦・クラスタ&グリッド計算の適用例について-, 情報処理学会誌 Vol.45 No.4, pp.372-376 (2004).
- [3] 神保進一: 最新マイクロプロセッサテクノロジー, 日経BP社, (1996).
- [4] 伊藤智: グリッドコンピューティングの技術動向, 情報処理学会誌 Vol.44 No.6, pp.576-580 (2003).
- [5] 溝口文雄: グリッドコンピューティング, 岩波書店, (2005).
- [6] 一条博: VPN 構築術, 工学社, (2004).
- [7] マルチメディア通信研究会編: インターネット RFC 事典, アスキー (1998).
- [8] 小早川知昭: IPsec 徹底入門, 翔泳社 (2002).
- [9] P. パチエコ (秋葉 博 訳): MPI 並列プログラミング, 培風館 (2001).
- [10] 戸川隼人: 共役勾配法, 教育出版 (1977).
- [11] 松田章, 加藤隆, 宮川佳夫, 河部徹, 長岡昭雅: 同報通信を用いた有限要素解析の並列高速化 (剛性方程式の並列高速求解), 日本機械学会論文集 65 巻 635 号 A 編, pp.1459-1466 (1999).
- [12] 松田章, 加藤隆, 宮川佳夫, 河部徹, 濱谷智哉: 非平衡木型通信方式を用いた剛性方程式の並列高速求解 (剛塑性有限要素解析の並列高速化), 日本機械学会論文集 66 巻 641 号 A 編, pp.20-25 (2000).