# 並列計算環境として利用可能な情報教育用PCシステムの構築

倉前宏行
大阪工業大学工学部技術マネジメント学科
kuramae@dim.oit.ac.jp

概要：夜間や休日といった情報教育用PC教室の休眠時間の計算機資源を有効活用するため、2003年より教室内のクライアントPCをクラスタ構成することにより、研究用の並列計算環境としても利用可能なシステムを構築してきた。本稿では、学科独自の教育用PCシステムを大学全体の基幹システムと連携した構成概要を紹介するとともに、並列計算環境としてのシステム性能についても述べる。

キーワード：OS混在環境、PCクラスタ、教育用システム、システム性能

## 1. 緒言

近年、システム工学分野の学部専門教育において、PCやネットワークといった情報機器を利用した教育機会が増大し、教育内容も多岐に渡ってきている。著者らはこれまで、こうした教育内容に柔軟に対応するため、1999年よりWindows系OS（Operating System）とPC UNIXのデュアルブート環境を提供する教育用システムを構築し、特に、OS混在環境における大量の学生ユーザ管理について、ユーザアカウントデータベースの統合管理方法に関する研究を進めてきた[(1)-(3)]。同時に、学部教育において、システムプログラミング、実験システム工学といった課題主導型の教育を実践してきた[(4)’(5)]。

教育用システムは、コンピュータに初めて触れるようなエンドユーザから、研究のためのシステムソフトウェア開発を行なう者まで多様なユーザが利用する。さらに、高速インターネット回線の普及によって、学生が自宅などの学外から、教育用システムにおいて提供している各種サービスやアプリケーションソフトウェアを利用したいといった要望も増えてきた。

そこで2003年以降、大学全体の情報システムを管理運用している情報センターと連携して、学科独自の教育用サブシステムを構築した。さらに、夜間や休日などのシステム休眠時間において、ネットワークを介して多数のPCクライアントを並列処理環境としても利用できるようにシステム構築を行なった。本システムは、2007年に機種更新を行ったものの、現在に至るまでの約6年間、システムの基本構成を変更することなく安定的に運用を継続している。本稿では、そのシステム概要について述べる。

## 2. 情報センターとの連携によるシステム構築と運用

### 2.1 連携ポリシー

学科独自の教育用システム構築は、情報センターのシステム更新とタイミングを合わせることにより、情報センターと協力し、電子メール、ユーザアカウント、個人ディスクといったユーザに提供するサービスを全学的視点で統一的に行なった。学科独自のシステムは、30台規模のPCクライアントを有する3つの演習室からなり、比較的小規模である。しかし、全学的には、約650台のPCクライアントを約7300名のユーザが利用する大規模なものである。よって、授業におけるサービスの一斉利用時のシステム負荷を想定し、安定的に動作するように導入前試験を繰り返し行いながら、システム設計や構成機種選定を行った。

学生ユーザにとって、Windows、Linuxそれぞれを利用する際のユーザ認証、電子メール利用時のユーザ認証に用いるユーザ名とパスワードは、学内で同一のものを利用するのが便利である。システム管理者にとっても、学生の入学と卒業によって、年に1度、全ユーザの約1/4が入れ替わるため、ユーザ管理はできる限り一元化して省力化を計ることが望ましい。

そこで、図1に示すように、学生ユーザの利用を（1）Windowsクライアントの利用、（2）Linuxクライアントの利用、（3）学外からのインターネット経由による利用、の3系統に絞り込み、このうち、パスワードのユーザ自身による変更、電子メールサービス、学外からのユーザ個人ディスクへのアクセスをWebブラウザによるHTTP（Hyper Text Transfer Protocol）によるアクセスに限定することとした。学外からアクセスする場合に

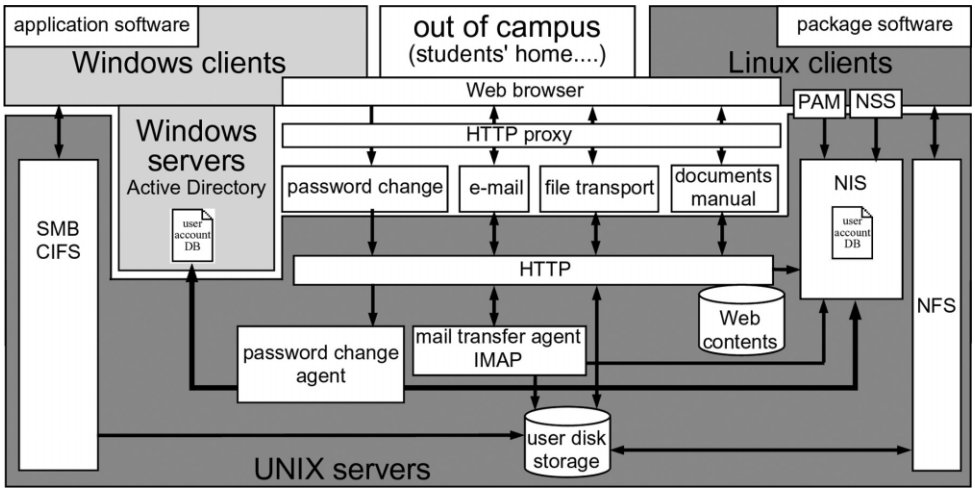


図1　OS 混在環境のシステム構成

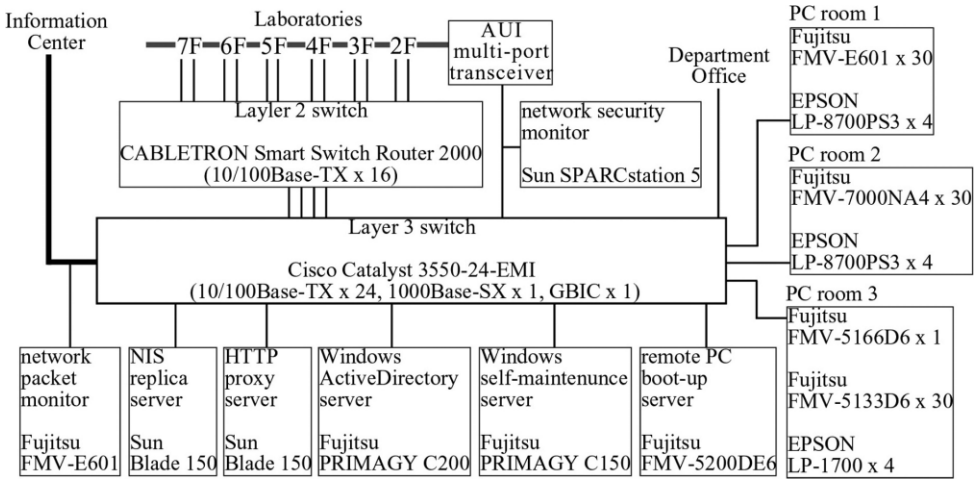


図2　学科サブシステムの機器構成（2003年当時）

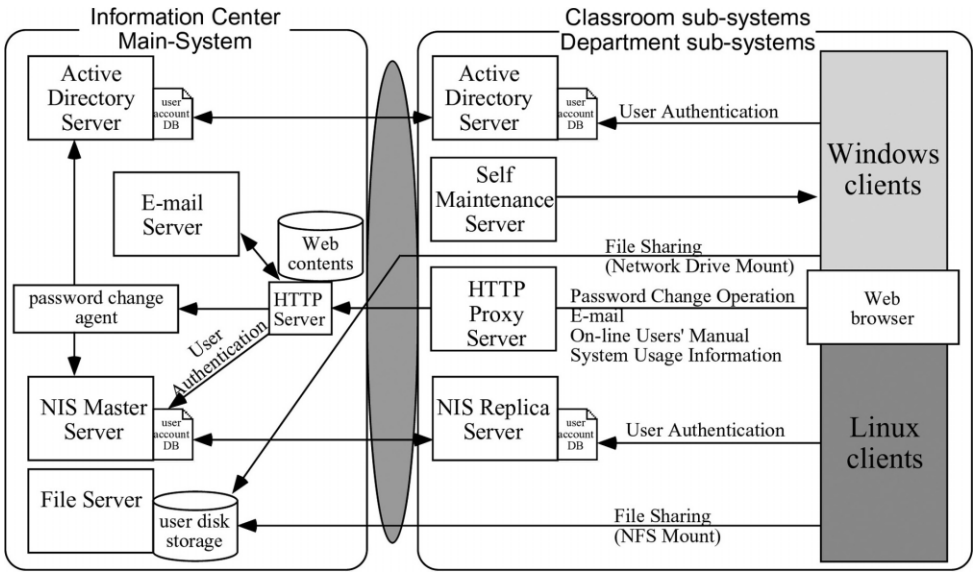


図3　情報センターシステムとの連携

は、セキュリティを考慮して、暗号化通信のHTTPS（Hypertext Transfer Protocol Security）を利用する。

以上の全学的統一ポリシーによって、管理対象のサービスや機器を最小限に絞り込ながら、ユーザにとってはWebブラウザのみでサービスを利用できるようなる。ここで利用する各種構成技術はいずれも開発から十分な時間を経て実績を積み重ねられているものばかりであり、大規模システムにおいて安定的に動作すると考えた。

### 2.2 ユーザ認証の統合

WindowsとUNIXとではパスワードの暗号化形式や認証方式が異なるため、ユーザアカウントデータベースを1つに完全統合して管理することは容易ではない。特に、ユーザ情報に関するネーミングサービスの応答性能は、OSの動作性能に大きな影響を及ぼすため、大規模システムにおいては、無理にユーザアカウントデータベースを統合すると、致命的なシステム障害を引き起こす[(1)]。

そこで、本システムにおいては、Windowsユーザアカウントは Microsoft社純正のActive Directoryを、Linuxユーザアカウントならびにメールサーバなどの UNIXサーバにおけるユーザ認証はNIS（Network Information Services）を用いて、それぞれ管理する。ユーザ登録時に、2つのユーザアカウントデータベースに同一の初期パスワードを設定しておく。ユーザ自身によるパスワード変更には、パスワード同期システム（富士通北陸システムズ社製、スルーPASS[(6)]）を利用する。Webブラウザを通してパスワード同期プログラム（パスワード変更を行なうラッパープログラム）にアクセスすることにより、2つのアカウントデータベースの内のパスワード情報を同時にそれぞれ修正変更する。これにより、ユーザからは完全にパスワードが一元管理されているかのように見える。

### 2.3 Webメール

ユーザが学内外を問わず、各種OS環境においてメールの送受信を行えるようにするため、Webメール（トランスウェア社製、Active!mail[(7)]）を導入した。本システムは、受信メールをIMAP（Internet Message Access Protocol）によって管理するとともに、Webブラウザで利用する。さらに、GUI画面のデザイン変更や転送メール設定、携帯電話からのアクセスもできる。

### 2.4 個人ディスク管理

学内共通のファイルサーバを導入し、Windowsのマイドキュメント、Linuxのホームディレクトリ、およびメール保存領域として、1ユーザあたり200MBの個人ディスク領域を割り当てた。WindowsにはCIFS（Common Internet File System）、LinuxにはNFS（Network File System）により全クライアントPCにディスク領域をマウントする。同様に、教員から学生への教材ファイル配布のためのディスク領域も用意した。

学外からインターネット通じて個人ディスク上のファイルにアクセスさせるため、Webブラウザを通したHTTPによってファイルのアップロード／ダウンロードが行えるインターフェースも用意した。

## 3. 学科サブシステムの構成

### 3.1 ネットワーク構成

図2に学科サブシステムの構成を示す。情報センターからGiga-bit光ケーブル（1000Base-SX）によって引き込まれたネットワークは、レイヤ3スイッチを通じて、教員居室など研究室系と演習教室系に分岐する。それぞれのネットワーク系統にはVLAN（Virtual LAN）を設定し、セグメント間のトラフィックを最小限に抑えるようにした。なお、サーバ群は全てレイヤ3スイッチへ直結している。

2003年当時は100Base-TXスイッチングネットワークで構成したが、2006年以降は1000Base-Tネットワークに増強し、サーバならびにPCクライアントも最新機種に随時更新している。

### 3.2 サーバ構成

電子メールやパスワード管理といった各種サービスは、情報センターで集中提供するが、ユーザ認証やユーザ情報の名前解決（ネーミングサービス）は、サブシステムごとに分散サーバを設置することで、システム負荷集中を回避した。本学科サブシステムにおいても、図3に示すように、Windowsサーバならびに、UNIXサーバを各2台設置した。Windowsサー

バは、Active Directoryバックアップサーバを通じてWindowsユーザアカウントを管理する。セルフメンテナンスサーバは、Windowsクライアントのファイルシステムを管理し、OS起動時にWindowsシステム領域（C：ドライブ）をシステム構築時の状態に完全初期化する[8]。また、ウィルス対策ソフトウェアのパターンファイルも、ここから自動配布する。

UNIXサーバでは、HTTPプロキシを提供し、演習室からの全てのHTTPアクセスを管理／キャッシュする。また、NISレプリカサーバを構築し、Linuxクライアントのユーザ認証とユーザ情報に関するネーミングサービスを提供する。

## 4. 並列計算環境

### 4.1 言語コンパイラと通信ライブラリ

クライアントPCのLinux環境は、ネットワーク接続された分散メモリ型並列計算環境として利用できるよう、さまざまな工夫を施した。言語コンパイラとして、Linux標準のGNUコンパイラに加え、Intel C++、Fortranコンパイラ（非商用版ライセンス）も追加導入した。

並列計算を行う際のメッセージパッシングライブラリとして、PVM（Parallel Virtual Machine)、ならびにMPI（Message Passing Interface）のフリー実装版MPICH[9] 1.2.59）（p4 デバイス）をインストールした。

### 4.2 遠隔電源制御とOS起動

本環境は、夜間や大学休業日といったPC演習設備の休眠時間に、PCクライアント資源を並列計算に利用することを想定している。しかし本学の場合、大学休業日や夜間にPC演習教室へ立ち入ることは許可されていない。したがって、研究室など立ち入りが許可された場所からネットワーク経由で演習教室内のPCクライアントの電源を投入し、Linuxを起動させる必要がある。そこで、遠隔電源制御機構を開発した。

PCクライアントのデュアルブート環境は、授業等の使用頻度に応じWindowsをデフォルトOSとしてあり、Linuxを起動させるためには、電源投入直後、OSブートローダ（GRUB）メニュー画面でLinuxを手動選択する必要がある。

図4に示すように、遠隔起動サーバは対象のPCクライアントに対して、電源投入指令のためのマジックパケットを送出する。PCクライアントは、PXE（Preboot eXecution Environment）機能によって、電源投入後、OSを起動する直前に、DHCP（Dynamic Host Configuration Protocol）サーバに対して、IPアドレスなどの問い合わせを行う。このとき、DHCPサーバのオ

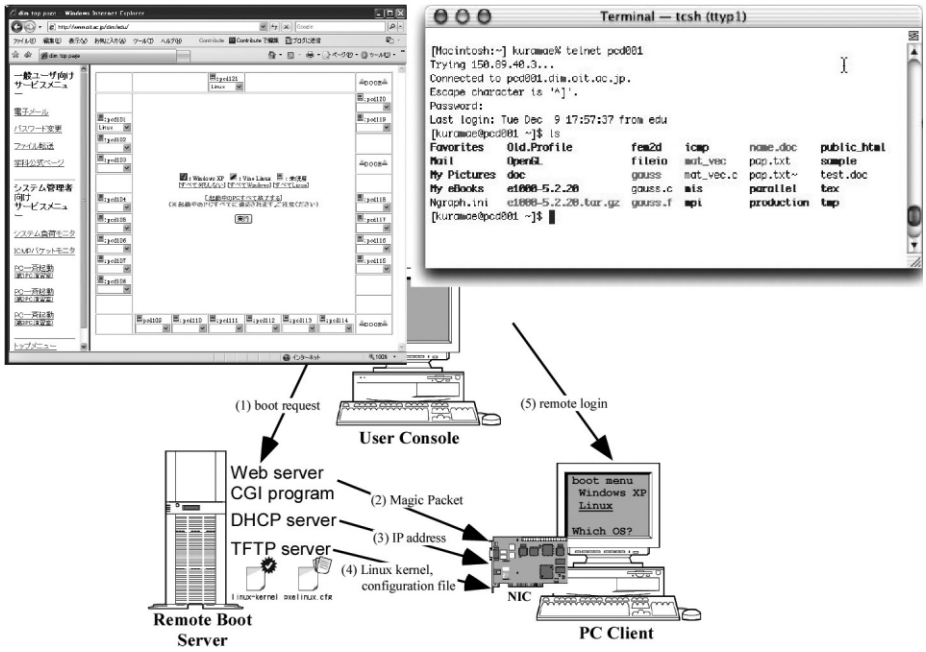


図4　Linuxリモートブート機構

表1　並列処理環境

| | Classroom PC system | | PC cluster system |
|---|---|---|---|
| | Fujitsu FMV-E601 | Fujitsu FMV-C5200 | VT64 Opteron WS 2000 |
| CPU | Intel Pentium4 2.4GHz | Intel Pentium4 3.2GHz | AMD Opteron244 1.8GHz |
| RAM | 512MB | 1GB | 2GB |
| OS | Redhat 9 | | SuSE Professional 9 |
| Network | Gigabit Ethernet 1000Base-T | | |

プション設定により、ディスクレスLinuxを実現するPXELinux10)を用いて、Linuxカーネルと設定ファイルをTFTP(Trivial File Transfer Protocol)によって転送する。これにより、PCクライアントは、GRUBを迂回してLinuxを自動起動することができる。なお、DHCPサーバにおいてPXELinux [10] の設定ファイルをローカルブートにしておけば、Windowsを起動させることもできる。そこで本機構は、図4に示すように、遠隔起動サーバに対して、起動させたいPCとOSをWebインターフェースを通してそれぞれ指定できるようにした。さらに、使用後には、同サーバを通じてPCのシャットダウンも行えるようにした。

### 4.3 計算ノードCPU　負荷モニタ

並列計算を行う際、計算ノード（PCクライアント）のCPU負荷状態をリアルタイムに把握する必要がある。そこで、計算ノード負荷モニタを開発した。負荷モニタサーバから、SNMP（Simple Network Management Protocol）を利用して、ネットワーク経由でPCクライアントのCPU負荷を問い合わせ、負荷推移を時系列的にグラフ表示する。本モニタのGUI開発には、Perl/Tkを用いた。

### 4.4 並列計算環境としての性能評価

PCクラスタによる並列計算においては、ネットワークの通信速度がボトルネックとなる。本システムを並列計算に利用した際の通信性能を実測し、並列計算専用のPCクラスタシステムとの比較を行った。表1に示すように、2種類のPCクライアントとPCクラスタとを比較した。

図5に実効ネットワーク転送速度の実測値を示す。1対1ブロック通信を行うMPI関数MPI_Send()、MPI_Recv()を用いてデータ送受信したときの転送速度を計測した。これより、PCクラスタ専用機と比較して最大で約2.6倍の性能差が確認される。これは、PCクラスタ専用機は大量のデータを高速に転送できるようNICデバイスドライバなどのチューニングが施されていることなどによるものと考えられる。

動的・非線形有限要素解析のように、計算ステップの途中で計算途中経過をファイルに書き出す並列プログラムは多い。本システムのLinux環境は、ユーザホームディレクトリを情報センターのファイルサーバにNFSで一元化しているため、ネットワークを通じたI/Oが並列計算実行時のボトルネックとなる可能性がある。そこで、Fortranプログラムによってファイルに書き出す際の時間をI/O時間を測定した。

1度に書き出すデータサイズを変化させたときの実効速度を図6に示す。同図より、ファイルサーバ上に書き出す場合には、ローカルディスクを用いる場合に比べて約7倍の時間を要している。したがって、ファイルI/Oを必要とする計算を行なう場合には、ローカルディスクを用いなければならない。そこで、Linux環境のローカルファイルシステム上に、計算作業専用のディレクトリを用意した。

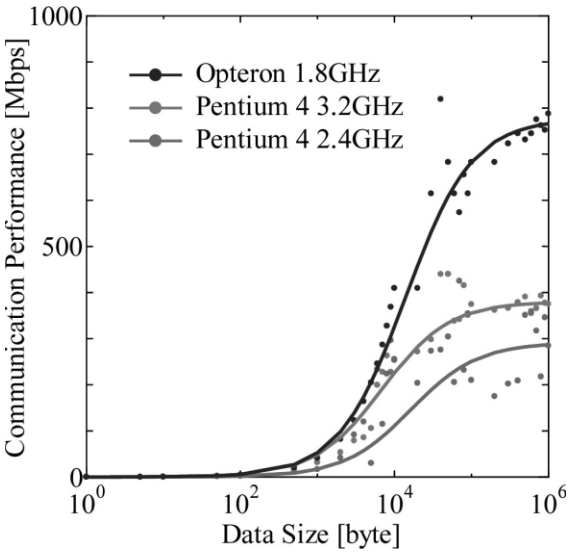


図5　ネットワーク通信性能

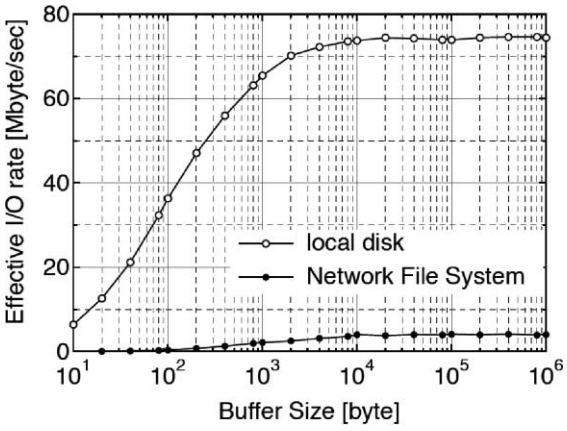

図6　ファイル書き込み速度

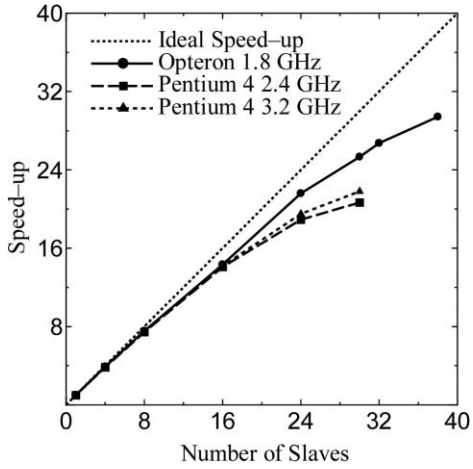

図7　並列性能の比較

実際に本システムにおいて、板成形問題の結晶均質化マルチスケール有限要素解析[11]を並列計算した場合のスピードアップ（逐次計算に対する並列計算の速度向上率）を、図7に示す。ネットワーク通信性能の差により、PCクラスタ専用機に比べて本システムの並列性能は低いものの、30台のPCを用いても70%を超える並列性能が得られており、休眠時間における計算機資源を有効に活用できる。

## 5.　結言

並列計算環境としての利用を念頭において、学科教育用PCシステムを構築した。情報センターとシステム連携をはかることで、教育システムとしての各種サービス提供を大学全体のシステムと共通化した。並列計算環境として利用するために、遠隔電源制御・OS起動機構、ならびにCPU負荷モニタを開発した。本システムを並列計算に利用する学生は、一部の研究室に配属された学部4年生や大学院生に限られるものの、有効に活用している。2008年より、VPN（Virtual Private Network）の運用がはじまった。遠隔電源制御・OS起動機構により、学生が自学自習や課題遂行のために自宅などからWindowsリモートデスクトップを使って利用することも技術的に可能となった。この運用方法について現在検討を進めているところである。

最後に、本システムの構築にあたり、大阪工業大学情報センター、富士通(株)、(株)富士通ビジネスシステムの関係各位に多大なご協力をいただいた。ここに、感謝の意を表わす。

## 参考文献

(1) Kuramae, H., Shimano, A., Kimura, A., Matsumoto, M. Furuno, Y. and Kamejima, K.,：User account management of PC cluster system for education using directory service, Proc. of World Multiconference on Systemics, Cybernetics and Informatics, Vol. VIII, pp. 68-73 （2001）.

(2) 倉前宏行、島野顕継、木村彰徳、松本政秀、古野良樹、亀島鉱二：複数OS利用のためのユーザアカウントデータベース統合、情報処理学会分散システム/インターネット運用技術シンポジウム2002 論文集、pp. 63-68（2002）。

(3) Kuramae, H., Shimano, A., Kimura, A., Matsumoto, M., Furuno, Y. and Kamejima, K.：Integration of user account database on multiple OS environment, Proc. of IEEE Region 10 Conference on Computers, Communications, Control And Power Engineering, Vol. I, pp. 176-179（2002）.

(4) 倉前宏行：情報教育用PCクラスタシステムの構築と経営工学科における計算力学教育、日本機械学会関西支部第23回NCP研究会・機械の強度と形態研究懇話会シンポジウム論文集、pp. 21-23（1999）。

(5) 倉前宏行、木村彰徳、中西真悟、下左近多喜男、上野精順、松本政秀、亀島鉱二：システム工学領域における実験演習科目の創設、私立大学情報教育協会第9回情報教育方法研究発表会予稿集、pp. 34-35（2001）。

(6) スルーPASS, http://jp.fujitsu.com/group/fjh/services/package/tpass/

(7) Active!mail, http://www.transware.co.jp/product/am/

(8) 瞬快 http://jp.fujitsu.com/group/shikoku/services/packages/shunkai/

(9) MPICH
http://www.mcs.anl.gov/mpi/mpich/
(10) PXELinux
http://syslinux.zytor.com/pxe.php
(11) Kuramae、H., Okada、K., Yamamoto、M., Tsuchimura, M., Sakamoto、H. and Nakamachi、E.：Parallel performance evaluation of multi-scale finite element analysis based on crystallographic homogenization method, Computational Plasticity, Owen, D.R.J., Onate, E. and Suarez, B. eds. CIMNE Barcelona、Part 1, pp. 622-625（2005）.