VMware と Dual-core PC を用いた PC cluster と PC 教室との並列共存システムの構築

Construction of the Computer System to coexist commodity PC clusters with PC rooms using VMware and Dual-core PCs

幸谷智紀*

Tomonori KOUYA*

Abstract: In recent years, most of middle scale scientific computations are executed on low-priced commodity PC clusters, which can be used as PC rooms for educational exercises. But these PCs in universities are always used for extracurricular activities, public relations or regional services even during vacation, so simulation programs for scientific computations which need long running time must be run in night or less idle days. Therefore it is desirable that these programs can be run during educational use on the same PCs without conflict of resources. In this paper, we report our experiments in progress to prove that both educational use and scientific computation can be run simultaneously by using virtual PC technologies like VMware.

1. 初めに

コンピューティング環境が大型計算機から PC へと移行して以来，多くのユーザが大規模な科学技術計算を行うために PC cluster を利用するようになった。特に，安価な Ethernet(10〜1000BASE) で市販の安価な PC を組み合わせた Commodity PC cluster は手軽に構築できる上，教育用 PC としても普通に使用できるため，大学等の教育研究機関では広く普及している。また，教育用に導入された実験演習用 PC を Campus Grid として利用することも行われるようになった [10]。

近年は学生募集のための広報を兼ねた高校生向け実験講座や，地域社会向けの講習会などで大学のコンピュータ・ネットワーク資源を利用する機会が増えている。加えて，在学生への教育は一層の充実を図ることが社会から求められており，平日，休日，長期休暇問わず，年間を通じてこの資源を隙間なく活用していく必要がある。この場合，主として Windows[11] 上でアプリケーションソフトを使用することが多い。

しかしながら，研究活動も従来以上の成果を求められるようになっており，おろそかにはできない。現在，Windows でも並列分散プログラムを動作するための環境が整いつつあるが [12]，多くの既存の科学技術計算ライブラリは UNIX 環境 (Solaris, Linux, *BSD 等) で利用されることが想定されており，前述の用途で PC が使用されていると共存ができない。従って，研究専用の大規模な PC cluster を設置できない教育機関においては，特に長い実行時間を要する大規模なシミュレーションが必須である研究活動を行うために，比較的少数の PE 数で構成される専用 PC cluster を細々と使うか，夜間を中心とした限られた時間帯を狙って間歇的に Job を流すか，いずれかの状況に甘んじなければならない。

しかし，PC に使用される CPU の性能は格段に進化しており，Commodity PC でも Core を複数搭載した Multi-core CPU が普通に利用されている。そこで，近年注目度が増した仮想 PC 技術を用いて，教育用の Windows 環境と，研究用の UNIX 環境との共存実験 (Fig.1 参照) を行ってみた。このためには幾つか方法や OS の組み合わせがあるが，我々は仮想化 PC 環境には VMware[5] を使用し，Fedora Core 5[13] を Host OS，Windows

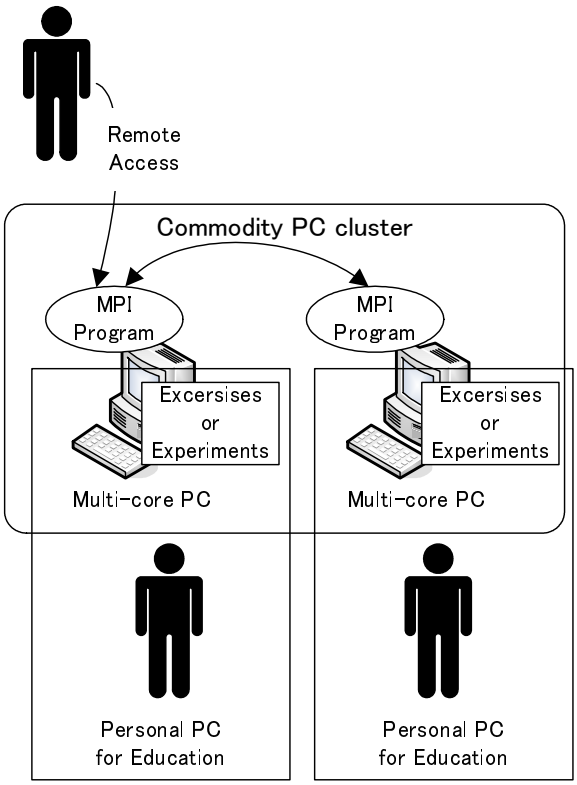

Fig. 1: 共存実験のゴール

2007 年 2 月 26 日 受理

*理工学部情報システム学科

XP Professional を Guest OS として実験を行っている。本稿ではこの実験の概要と途中結果について報告する。

2. VMware と Multi-core CPU による共存環境

VMware は仮想 PC 技術をいち早く取り入れて世に出された商用ソフトウェアである。2006 年 9 月現在，データセンター用の Infractructure, 開発者用の Player/Server/Workstation がある。今回は Fig.2 のように，ベースとなる Host OS の上に仮想 PC 環境を構築する後者のみを視野に入れる。

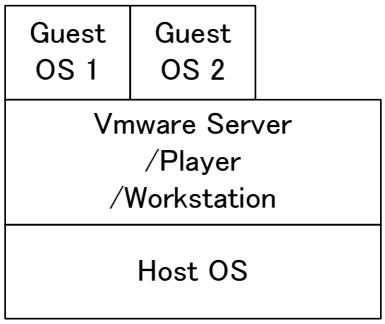

Fig. 2: VMware Workstation/Server/Player のソフトウェア構成

後者のうち，現在では Player と Server が無料で使用できるようになっている。そのうち Server は単独で Geust OS 環境を複数の仮想 PC 上に構築でき，ネットワーク環境構築実験 [6] 等にも転用できるので，今回はこれを使うことにする。

古くから UNIX 雑誌で取り上げられてきただけあって，VMware の仮想 PC 環境の出来の良さは定評がある。しかし個人的に試してみた限りでは，Pentium 4 以前の CPU を積んだ PC ではいささか動作が重いという印象があった。しかし，Pentium 4 以降はメモリさえ十分に積んでいればかなり実用になると感じ，これが Multi-core CPU であれば，複数の OS を同時実行しても，多少のもたつきはあっても実用には問題のないレベルになると確信した。

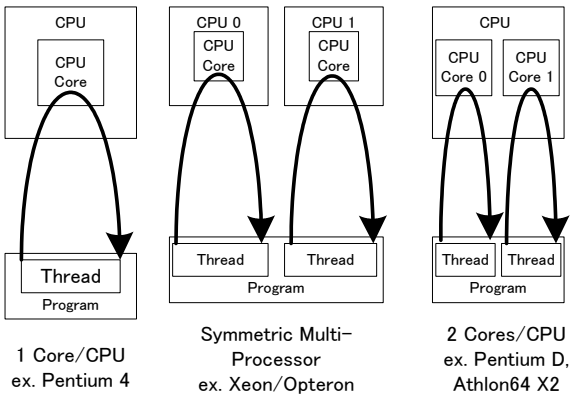

Fig. 3: Single core, SMP と Multi-core CPU

CPU 内部の動作 clock が 1GHz を越えて以来，clock 数の向上率は鈍化している。その原因は電力消費の増大にあり，これを解決するために，CPU core を複数搭載する Multi-core 化が 2005 年から本格的に始まっている。Multi-core CPU は実質的には SMP と同じレベルの高い並列処理性能が期待できる (Fig.3) が，単一のソフトウェアがその性能をフルに使いきるためには Multi-thread 化するなどの書き換えが必須であり，現状多くのアプリケーションは Single Thread 動作になっていると思われる。これが逆に，Host OS での Multi-core の使い切りを防いでおり，我々が目的とする共存実験には都合が良い。

よって，“Host OS+ 研究用プログラム”と，“Guest OS(+ 教育用アプリケーション)”とで複数の core を分け合えば，CPU 資源の食い合いは防止できる。しかし，それ以外の資源についてはそれぞれの要求が衝突しないよう，留意する必要がある。(Fig.4) 具体的には以下の 3 条件となろう。

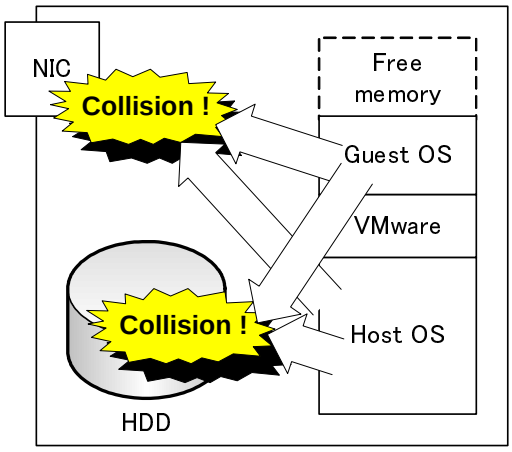

Fig. 4: 共存のための 3 条件

1. Host OS(+ 研究用プログラム) + VMware + Guest OS(+ 教育用アプリケーション) を全て合計しても，搭載してある RAM 容量内に余裕を持って収まる
2. ネットワーク資源を大量に使用しない (片方のみなら O.K.)
3. 外部ストレージ (HDD, 光ドライブ等) 資源を大量に使用しない

1 の条件を満足しなければ当然仮想記憶を食いつぶし，HDD への大量の Swap が発生するため，結果として 3 の条件をも満足しなくなってしまう。従って，1 は必須条件であると言える。

これらの条件を満たす限り，両者の共存は十分可能であると予測できる。

3. 共存実験結果

共存実験には次の計算機環境 (PentiumD) を用いた。Pentim D は Fig.3 に示すように 2 つの CPU core(Dual core) を搭載しているため，前述の通り，Fig.5 のような共存が可能であると予測できる。

今回は，同じ Pentium D を積んだマシンを 4 台 (cs-room443-d01～d04) 揃え，MPI とライフライン向けに二つの GbE を用意した Commidity PC cluster としても活用している下記の PC 構成を使用する。

PentiumD Intel Pentium D 820 (2.8GHz), DDR2 4GB RAM, Fedora Core 5 x86_64

VMware VMware Server 1.0.0 Build 28343 (for Linux)

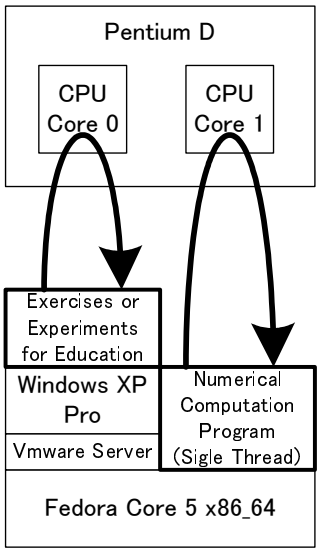

Fig. 5: Pentium D における共存実験

Windows　Microsoft Windows XP Professional SP2 (32bit)

仮想 PC には 1 CPU, 1024MB RAM を割り振ってある。また，VMware による仮想 PC 環境に Windows XP とアプリケーションを搭載するには，少なくとも 10GB 以上のディスク領域を確保する必要があるため，今回は 20GB 確保してある。このサイズの disk image ファイルを，例えば NFS で共有し，しかも実用に耐えるだけの高速転送を行うには，現状の GbE ではかなり難しいと思われる。Node ごとにそれぞれローカル HDD に disk image を置くとしても，現状では 20GB の local copy だけでも 20 分〜30 分かかってしまうため，全て各マシンのローカル HDD 上に確保した。

これらの環境を使用し，まず，通常のアプリケーションの使い勝手と資源使用状況を調査する。次に，CPU core をフルに使用する処理を両者の環境で行い，互いの速度低減の状況を調査する。

なお，比較検討用に次の Pentium4 マシンも使用する。

Pentium4　Intel Pentium 4 2.8cGHz, 1GB RAM, Windows 2000 Professional SP5

3.1　MuPAD と OpenOffice を用いた高大一貫夏季実験講座

2006 年 8 月 24 日 (木) 13:00 から翌 25 日 (金) 16:00 まで，上記 Pentium D マシンを用いて高大一貫教育の夏期実験講座を行った。内容は，数式処理ソフトウェア (MuPAD Light 2.5.3) を使って与えられた問題を解き，その結果を OpenOffice Write 2.0 を使ってレポートに仕上げる，というものである [1]。MuPAD と OpenOffice は VMware Server 上の Windows XP Pro で動作させた。その際の CPU 負荷を Fig.6 に，Free memory 量を Fig.7 に示す。

cs-room443-d01 と d03 において，火曜日と水曜日の CPU 負荷が高いのは，試しに MPI 並列プログラムを実行したためであり，この実験講座とは関係ない。VMware は前日の水曜日午後から動作させ，翌週月曜日の午前中に停止，その後全マシンを再起動させている。

実験期間中の木，金曜日の CPU 負荷は殆ど低い状態であるが，d03，d04 については規則的に短時間ではあるが負荷が

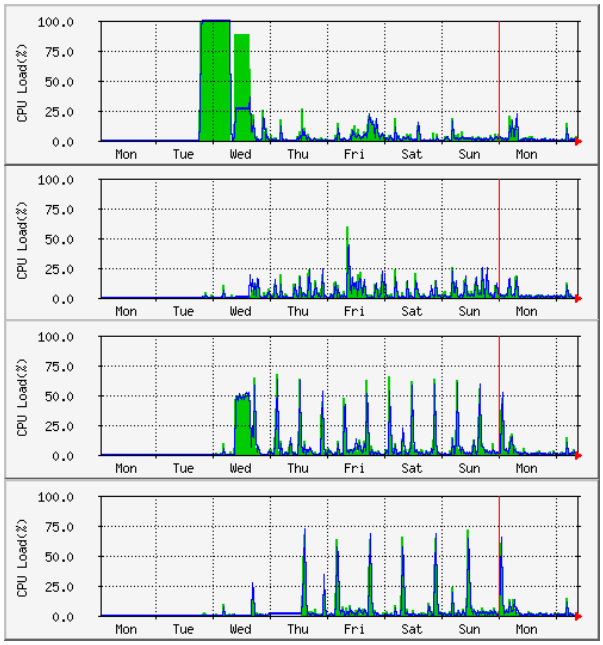

Fig. 6: CPU load average: 上段から cs-room443-d01, d02, d03, d04

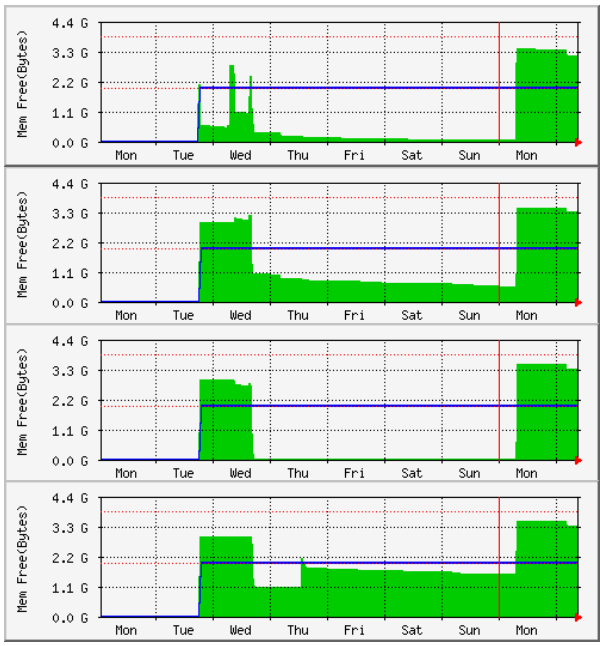

Fig. 7: Free memory: 上段から cs-room443-d01, d02, d03, d04

50%から 75%に上がる現状が起きている。これは再起動後には起きていないことから，VMware と何らかの Cron との競合によるものと思われる。

また，Free memory 量は VMware 動作中は低下していることが分かる。特に d01 と d03 が殆ど使い切った状態に見えるが，実験自体には全く支障をきたさなかったため，実際の状態は d02, d04 と同じであると思われる。この差異の原因は何かは不明である。

以上のように，CPU 負荷と Free memory 量の推移に問題は見られるものの，高大一貫実験自体は滞りなく終了し，受講生も使用している環境が VMware 上のものとも気がつかなったぐらい体感速度に問題が発生することはなかった。よって，我々の提案する環境では，今回の実験内容程度のものは十分に対応可能であると言える。

3.2 Let't Edit 2 for School による Movie encoding ベンチマーク

次に，CPU 資源を使い切る二つのプログラムを同時に実行し，速度低下の状況を見ることにする。

今回は，教育用アプリケーションとして，Let's EDIT 2 for School[16] を動作させ，40.4MB(2 分 45 秒) の MPEG1 ファイルに Movie encoding を行い，48KHz のサンプリングレートで WMV ファイル (生成後は 6.93MB) に変換する。同時に，研究用プログラムとして，ATLAS[8] の dgemm 関数による実正方行列積 (128〜5104 次元) を繰り返し実行し，GFLOPS 値を得る。これは IEEE754 倍精度計算主体の処理である。比較のために，整数演算主体の多倍長浮動小数点数 (仮数部 8192bits, 10 進 2466 桁相当) を用いて同じく実正方行列を行ってみた。

その結果を Table 1 に示す。全て 5 回実行した結果の平均値である。

Table 1: Movie encoding 実験結果

Env.	Elapsed sec.
Pentium4	75.5
PentiumD	69.8
& VMware	147.0
& ATLAS	155.4
& BNCpack	158.8

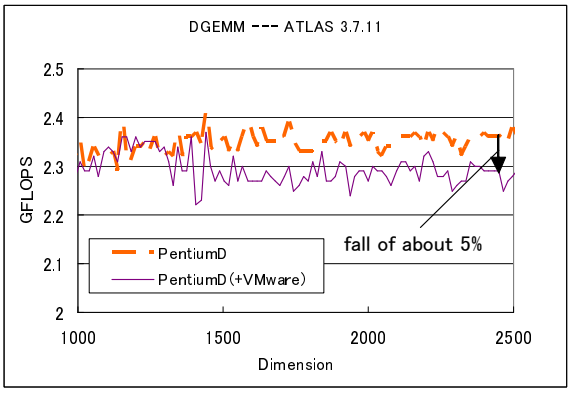

Fig. 8: ATLAS 行列積の性能低下

素の Pentium D(Windows XP を直接インストール) では Pentium4 に比べて 8.3%の速度向上が見られるが，VMware の仮想 PC 上では 47.5%の速度しか得られず，ほぼ速度は半減することが分かる。しかし，同時に Host OS 上で ATLAS 行列積や多倍長行列積を実行してもそれぞれ 2.6%, 3.5%しか速度は低下しない。同時に動作している ATLAS 行列積も約 5%程度の速度低下に留まっている (Fig.8)。

従って，Pentium D マシンにおいては，仮想 PC 上では処理速度が半分に落ちるものの，Host OS 下のプログラムとは互いにせいぜい 3%から 5%程度の性能低下に留まり，十分共存できることが判明した。

3.3 MPICH を用いた PC cluster 構築実験

以上の結果より，Multi-core CPU マシン上においては，VMware で適切に CPU 数を割り振ることにより，core 数分の Node を持つ PC cluster を構築しても，それに応じたパフォーマンスが得られると予想される。今回使用した Pentium D マシン上においては 1CPU/Node を割り振ることで，2 Node の PC cluster を構築できることになる。

そこで，Pentium III/IV Cluster 用に作成した MPI Cluster 構築マニュアル [2)3)] に基づいて，Vine Linux 3.2[14] を VMware 上の仮想マシンにインストールして MPI PC cluster を構築し，ベンチマークテストを行った。

まず，BNCpack[9] による多倍長正方行列積 (1024 次元) の実行時間を計測する。その結果を Fig.9 に示す。

この結果，仮想 PC cluster の性能は，同じ動作周波数である Single-core の Pentium 4 cluster より格段に良く，Pentium D ネイティブの性能に近いことが確認できた。

次に，VMware 上に構築した仮想 PC cluster 2 node 間の MPI 性能を，NetPIPE 3.6.2[15] で計測する。

仮想 PC cluster のネットワークは Pentium D マシン内部に構築されるため，memcpy 転送能力程度のパフォーマンスが出ても不思議ではない。しかし実際に計測してみると，ちょうど GbE(1000BASE-T) を搭載した Pentium 4 cluster 程度の性能しか出ていない。これが VMnet による制限なのかどうかは今のところ不明である。しかし，逆に言えば，ネットワーク能力に関しても Pentium 4 cluster と同程度であるため，CPU 性能に応じた並列性能を得ることできている訳である。

以上のように，仮想 PC clsuter においても CPU core 数に応じたパフォーマンスが得られることが判明したので，3 年生対象の情報セミナー 2 において，PC cluster 構築実習も行った。そのためのマニュアル (Fig.11) も作成し，Web 上で公開した [4)]。

4. まとめと今後の課題

以上の教育用の運用及び負荷実験と仮想 VM による PC cluster のベンチマークの結果，提案のシステムでは PC 資源の衝突が起きなければ問題なく運用できることが確認できた。

今後の課題としては次の 2 点が挙げられる。

1. 今回の実験はまだ Commodity PC cluster の 1 node だけ用いたものであるが，ネットワーク資源の食い合いさえなければ，MPI 並列処理プログラムも教育用環境と共存して実行できることが期待できる。ベンチマークテストを重ね，実証していきたい。
2. 今回構築した教育用の仮想 PC 環境は，全てローカルハードディスク上に構築されたものであるため，一度環境を構築してしまうと使用すべき PC が固定されてしまうという問題がある。しかし，現在安価に提供され

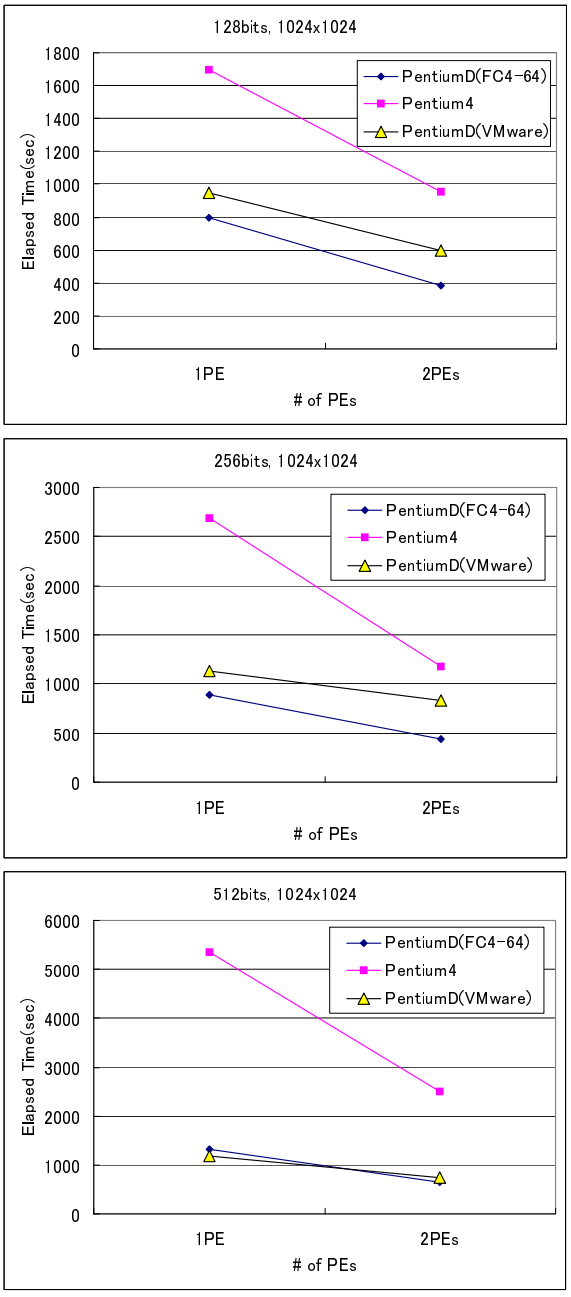

Fig. 9: 多倍長正方行列積の実行時間 (上から 128, 256, 512bits 計算)

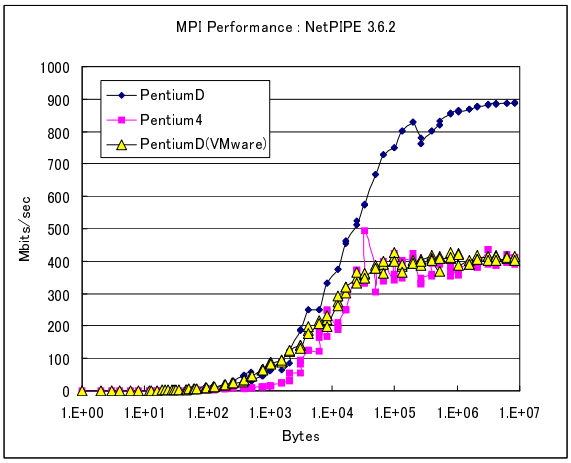

Fig. 10: VMware 上のネットワーク性能

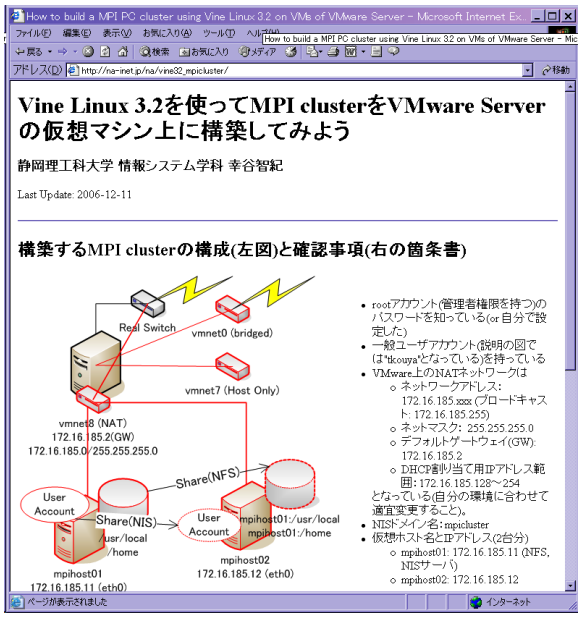

Fig. 11: VMware による PC cluster 構築マニュアル

ている GbE(1000BASE) では数十 GB もの仮想ディスクをネットワーク上で共有することは，パフォーマンスの点で問題がある。仮想ディスクをどのように共有していくか，という点も解決すべき今後の課題である。

謝辞

本研究は静岡理工科大学教育開発費の援助を得て行われた。また，夏期実験講座では千葉県立八街高等学校教諭・角谷悟氏の助力を，ベンチマークテストのデータ収集には静岡理工科大学 4 年生・芥田雄介君の助力を得た。厚く御礼申し上げる。

参考文献

1) 幸谷智紀・角谷悟, 数式処理ソフト MuPAD と OpenOffice を用いた低コストな高大一貫実験講座について, To appear.
2) 幸谷智紀, Vine Linux による PC cluster の構築, http://na-inet.jp/na/mpipc.pdf
3) 幸谷智紀, Vine Linux による PC cluster の構築, http://na-inet.jp/na/mpipc2.pdf
4) 幸谷智紀, Vine Linux 3.2 を使って MPI cluster を VMware Server の仮想マシン上に構築してみよう, http://na-inet.jp/na/vine32_mpicluster/
5) VMware, http://www.vmware.com/
6) 大江将史, VMware で UNIX, UNIX magazine, 2001 年 1 月号, pp.51-65.
7) Let's Edit 2 for School, http://www.canopus.co.jp/catalog/letsedit2/letsedit2_school_index.htm
8) Automatically Tuned Linear Algebra Software, http://math-atlas.sourceforge.net/
9) BNCpack, http://na-inet.jp/na/bnc/
10) 広島大学情報メディア教育研究センター, http://www.media.hiroshima-u.ac.jp/
11) Microsoft Windows, http://www.microsoft.com/japan/windows/
12) Windows Compute Cluster Server, http://www.microsoft.com/japan/windowsserver2003/ccs/
13) Fedora, http://fedora.redhat.com/
14) Vine Linux, http://www.vinelinux.org/
15) NetPIPE, http://www.scl.ameslab.gov/netpipe/
16) Let's EDIT 2 for School, http://www.canopus.co.jp/catalog/letsedit2/letsedit2_school_index.htm