

ネットワークブートシステムにおける ディスク性能の影響とその評価

宅間 広大^{†1} 榎田 秀夫^{†2}

運用管理のコスト削減や情報漏洩対策の一つとして、企業や教育機関でネットワークブートを用いた情報端末の運用が増えてきている。ネットワークブートを行う場合、通常、サーバが複数のクライアントに対する処理を行う。そのため、サーバのネットワーク性能とディスク性能がボトルネックとなると考えられる。本稿では、このうちディスクの実効性能を向上させる手段として、通常のハードディスクではなく、より高性能とされる、DRAMをベースとした製品であるGIGABYTE社のi-RAM、フラッシュメモリを使用したSSD、メインメモリを使ったRAMディスクをOSイメージの供給領域として使用することにより、どのような性能向上が見込めるかを明かにする。性能評価の指標として、クライアントPCの起動にかかる時間を使用し、自動的に起動時間を収集する仕組みを実装して、実機による評価を行った。その結果、i-RAMやメインメモリを使用した場合に、最大30%以上の性能向上が見られた。

Evaluation of efficiency of disk performance on Network Boot System

HIRO TAKUMA^{†1} and HIDEO MASUDA^{†2}

In companies and universities, network boot system is widely used for client computer system. It is good solution for reducing the total cost of ownership and security risks. The bottlenecks of the network boot server are local hard disk and network device because many client PCs access the boot server at once. In this paper, we show the impact evaluations of different types of disks; normal Hard disk drive, i-RAM based on DRAM (a product of GIGABYTE Inc), SSD (solid state disk) based on flash memory and RAMDISK based on main memory. In our evaluations, we measure the client's boot up time in various number of PCs. We also developed the application of evaluating the boot up time automatically in the real PC system. As the result, we show that i-RAM and RAMDISK is 30% or more faster than normal HDD.

1. はじめに

近年、企業や教育機関において、多数のPCを使うようになってきている。多くのPCが利用されるようになると、運用管理にコストがかかる。また最近では、企業などでは情報の漏洩が問題となっている。これらの問題を解決するものとして、ネットワークブートシステムが使われるようになってきている。例えば、東京大学情報基盤センターではApple社のNetboot¹⁾が使われている。その他のよく利用されるものとしてArdence社のArdence²⁾がある。

ネットワークブートシステムにおいて、通常、多数

のクライアントに対してブートサーバが処理を行うため、システム全体の性能はブートサーバの性能に左右される。そのため、システムの性能の向上には、ブートサーバの性能を向上させる必要がある。ブートサーバに対してかかる負荷として、ネットワークカード(NIC)およびディスクへの負荷が挙げられる。

本稿では、ディスク性能に着目し、ランダムアクセス性能などの実効性能の高いディスクを使用することによる、性能向上の効果について述べる。DRAMをハードディスクと同じように扱うGIGABYTE社のi-RAM³⁾、フラッシュメモリを使用したSSD、メインメモリを使ったRAMディスクを、クライアントOSの供給領域として使用した場合、ネットワークブートシステムの性能がどのように向上するか評価する。

2節でネットワークブートシステムの構成について説明し、3節で性能向上法として本稿において提案する手法を述べる。そして、4節で評価および5節で考察を述べる。

^{†1} 京都工芸繊維大学大学院 工芸科学研究科 情報工学専攻
Graduate School of Information Science, Kyoto Institute of Technology

^{†2} 京都工芸繊維大学 情報科学センター
Center for Information Science, Kyoto Institute of Technology

表 1 ブートサーバのスペック

CPU	AMD Athlon64 Processor 3500+(2.2GHz)
M/B	HP Proliant ML 115 G5
メモリ	DDR2-6400 8GBytes
NIC	VIA Technologies Gigabit Ethernet VT6120
ディスク	Westan Degital WD3200AAKS (SATA150), GIGABYTE i-RAM (SATA150) SSD STT-FTM64GL25H (MLC) (SATA150), RAM (メインメモリ)

2. ネットワークブートシステム

本節ではまずネットワークブートシステムの構成に関する要素技術について述べ、次に、今回構成したシステムの構成について述べる。

2.1 ネットワークブートシステムの構成

Linux を用いたネットワークブートシステムでは、ブートサーバのディスク上に置かれている kernel イメージ (vmlinuz)、初期ラムディスク (initrd)、およびルートファイルシステムをネットワーク経由で取得し、一般的にクライアント PC のローカルディスクを用いずに動作するシステムである。

今回のシステム構築にはクライアント PC の起動には PXE (Pre eXecution Environment)⁴⁾を用いた。また、PXE を使用するブートローダとして、syslinux の一部である PXELinux を用いた。

2.2 ブートサーバの構成

本節では、本稿において構成したブートサーバについて述べる。ブートサーバのスペックを表 1 に示す。OS は Ubuntu Linux 9.04 を用いた。ブートサーバとして機能させるために次のソフトウェア群を用いた。

- DHCP サーバ:
dhcp3-server; Version: 3.1.1-5ubuntu8
- TFTP サーバ:
tftp-hpa; Version: 0.48-2.3ubuntu1
- NFS サーバ:
nfs-kernel-server; Version: 1:1.1.4-1ubuntu1

また、NFS のバージョンは NFSv3 を使い、仕様プロトコルは TCP、オプションとして nolock、rsizе=8129, wsize=8129 を設定した。

3. ブートサーバの高性能化

本節ではブートサーバの性能向上法について述べる。

ネットワークブートシステムでは、クライアント PC は OS イメージなどのデータをネットワーク越しに取得し起動する。その性質上、ブートサーバの性能がネットワークブートシステム全体の性能にそのまま影響を与える。このシステム全体の性能向上のためには、ブートサーバの性能を向上させる必要があると言えらる。

3.1 ボトルネックの解析

クライアント PC の起動は DHCP サーバへの要求から始まる。その後、TFTP サーバから vmlinuz および initrd を取得し、クライアント OS のブートプロセスが始まる。次に、NFS サーバ上にあるルートファイルシステムをマウントし、init スクリプト群を実行していく。

このスクリプト群やそれによって起動されるデーモンの実行に必要なファイルはすべて NFS で取得される。したがって、多数のクライアント PC が同時に起動すると、ブートサーバの NIC への集中とディスクアクセスの集中が問題になると考えられる。

例えば、1000base-T の NIC を使用したとすると、最大でおよそ 125MBytes/sec 程度の速度でデータの通信が行われる。また、SATA150 のハードディスクの理論上限値はおおよそ 150MBytes/sec である。しかし、今回用いた HDD のスペック⁵⁾を見ると、最大連続転送速度を 70MBytes/sec としている。このようにシーケンシャルアクセスであっても、今回のように HDD のキャッシュ (用いた HDD では 16MBytes) に乗りきらないような大きなデータを読み出す場合においては、高々 70MBytes/sec 程度の速度である。また、ハードディスクの物理的な構造上、ランダムアクセスに対してそこまでの性能が出ることはなく、一般的にこれよりさらに下回ることが知られている。

そこで、ランダムアクセス時のディスク装置の性能を向上させることで、ネットワークブートシステムの性能がどのように向上するかを調べる。

3.2 ディスク性能の向上方法

前節で述べたように、ブートサーバのディスクの I/O 性能がボトルネックとなりうる。ランダムアクセス時でも性能の落ちにくいディスク装置として、i-RAM や SSD などが知られている。

i-RAM は通常メインメモリとして使用されている DDR メモリを、あたかもハードディスクのように見せかける製品である。SSD はフラッシュメモリを用いており、これらのディスク製品はランダムアクセスに対しても十分高速に動作するとされている。これらの製品は SATA 規格で PC に接続されるものを用いている。i-RAM において、文献³⁾によれば、同製品は多くのアクセスパターンに対して SATA 規格の理論値である 150MBytes/sec に近い性能を発揮する。SSD において、文献⁶⁾によれば、同製品はシーケンシャルリード時で最大 150MBytes/sec となっている。また、メインメモリを使った RAM ディスクでは、SATA 規格の制約を受けないため、より高速であると考えられる。

4. 計測結果

本節では起動時間、ネットワーク負荷、および、ディ

スク I/O 負荷の評価方法について述べ、今回計測した結果を示す。

4.1 システム構成

文献⁷⁾を参考に、unionfs を用いた共通化を図ったルートファイルシステムを構成した。クライアント OS として Vine Linux 4.2 を使用した。

4.2 クライアント PC の起動時間

今回の計測では、Wake On LAN (WOL) を用い、ほぼ同時にクライアント PC を立ち上げた。開始時刻は WOL のマジックパケットを送り始めた時刻とし、完了時刻は、すべてのクライアント PC において、自動ログインを実施した上で、ウェブブラウザ (Firefox) の自動起動が完了した時刻とした。

ブラウザ起動完了時刻は、自動ログイン専用のユーザがログインした後自動実行されるスクリプトにより Firefox を立ち上げ、ホームページをロードし終えた段階で Javascript によりポップアップウィンドウを出すようにし、そのポップアップウィンドウのタイトルが xwininfo コマンドを用いて取得できた時刻とした。

ディスク、同時起動するクライアントの台数の組み合わせそれぞれにつき 5 回計測した。計測結果のうち上端と下端をのぞいた 3 値の平均値を表 2 に示す。これをまとめたグラフを図 1 に示す。

4.3 ネットワーク負荷

/proc/net/dev を 1 秒毎に収集し、受信および送信されたバイト数 (速度) を計測した。次に示すグラフ (図 2, 3) は、クライアント PC を 22 台同時起動させた場合のある 1 回の計測結果である。

4.4 ディスク I/O 負荷

/proc/diskstats を 1 秒毎に収集し、1 秒あたりの読み込みおよび書き込みのバイト数 (速度) を計測した。次に示すグラフ (図 4, 5) は、クライアント PC

を 22 台同時起動させた場合のある 1 回の計測結果である。

なお、メインメモリを RAM ディスクとして用いた場合、/dev/diskstats ではディスク I/O が取得できなかった。そのため、RAM ディスクを用いた場合のグラフを省いている。

5. 考 察

5.1 クライアント PC の起動時間

表 2 から、16, 22 台の同時起動時には、i-RAM を用いた場合で、それぞれ 16, 26% 程度の時間の短縮がみられ、また RAM を用いた場合には、37, 56% 程度の短縮がみられる。また、1~8 台の場合にも、RAM を用いた場合には 16~27% 程度の短縮がみられた。i-RAM についても、ほぼ同様のことが言え、1, 4 台の場合で 11, 13% 程度の短縮がみられる。しかし、8 台の場合には逆に HDD よりも起動に時間がかかっている。

5.2 クライアント PC の挙動解析

図 3 の HDD のネットワーク (送信) についてのグラフを見る。起動開始から 20 秒後のピークでは、vmlinuz と initrd のロードがなされており、30 から 75 秒程度の間にはローダブルモジュールが読まれている。その後、100 から 160 秒にかけて、デーモンの起動を行っており、これらの実行に必要なファイルを読んでいると考えられる。その後の 160 から 180 秒にかけて、デスクトップ環境を実行し、Firefox を実行している。そして、180 秒程度で起動が完了している。

5.3 ネットワーク負荷

図 2 から、受信においてはどのタイミングに置いても、高々 20MBytes/sec 程度の性能しか出ていないことがわかる。起動時刻から 60 秒程度経過した頃に i-RAM および RAM において、ピークがみられる。これは、ディスクの書き込み (図 5) におけるピークと一致する。

ブートサーバのネットワークでの受信として考えることができるものは、NFS サーバへのファイルの要求と、/etc や /var への設定ファイルおよびログファイルの書き込みである。

図 3 より、送信においても高々 50MBytes/sec の性能しかでていないことがわかる。理論上限値は 125MBytes/sec に対して 50% 未満の負荷であり、22 台程度の規模ではネットワーク負荷はまだボトルネックとなっていないと考えられる。

また、送信においてもそれぞれのディスクを使用した場合に違いが出ている。この違いはデーモンを実行している 100 秒以後の大きな山に顕著に見られる。

起動後 20 秒のカーネル読み込みから、ローダブルモジュールの読み込みのあたりでは、ある程度グラフは重なっている。これは、ある程度同じファイルを読み出していることにより、この時点ではディスクの性

クライアント台数	1	4	8	16	22
HDD	88	97	114	158	181
iRAM	78	86	116	134	158
SSD	86	89	104	161	214
RAM	74	82	89	115	132

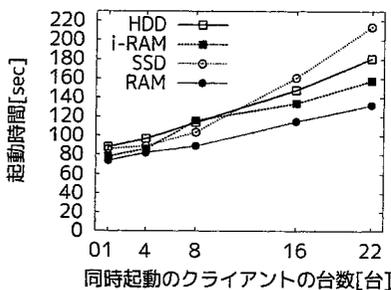


図 1 起動時間の計測結果

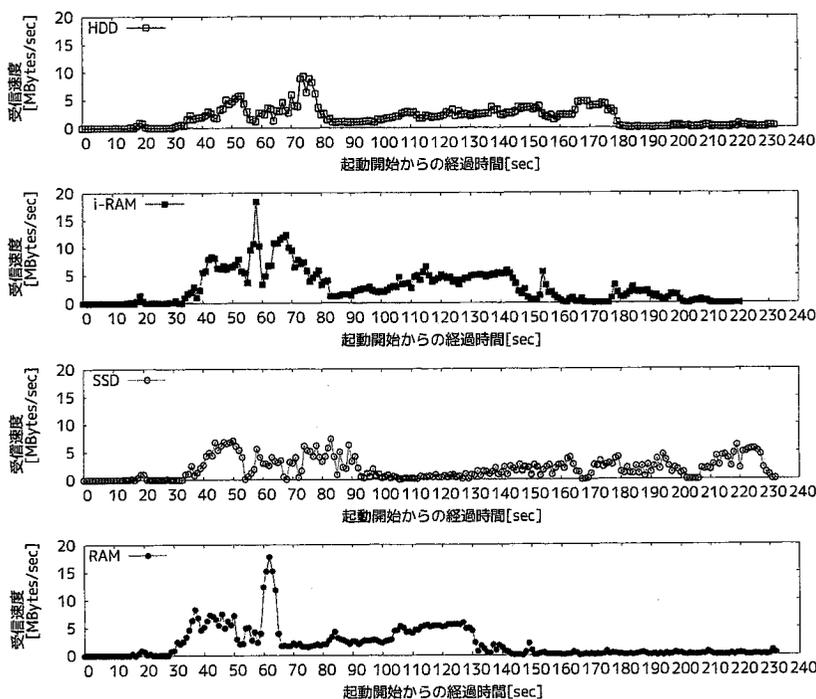


図2 クライアント 22 台同時起動時, ある 1 回のネットワーク (受信) の計測結果

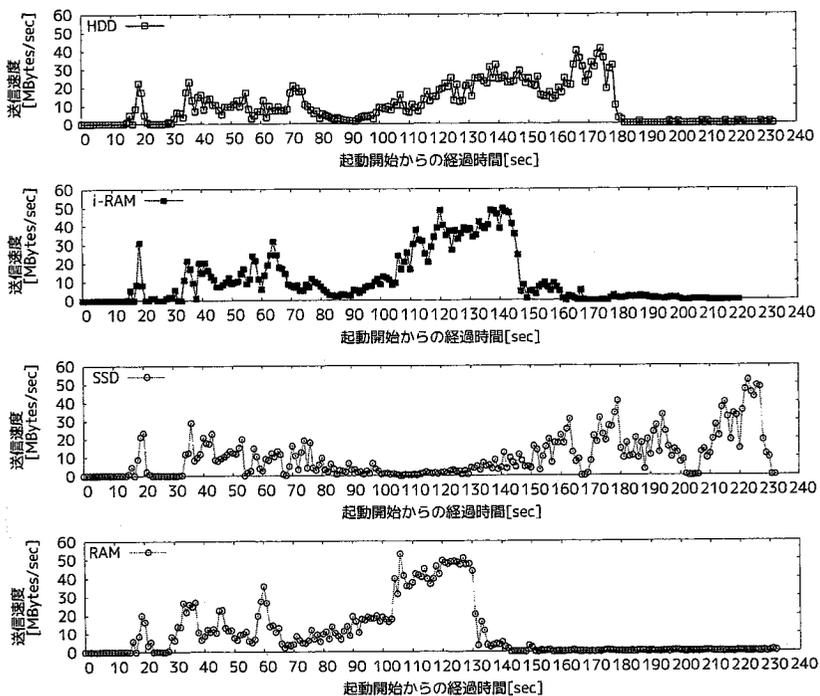


図3 クライアント 22 台同時起動時, ある 1 回のネットワーク (送信) の計測結果

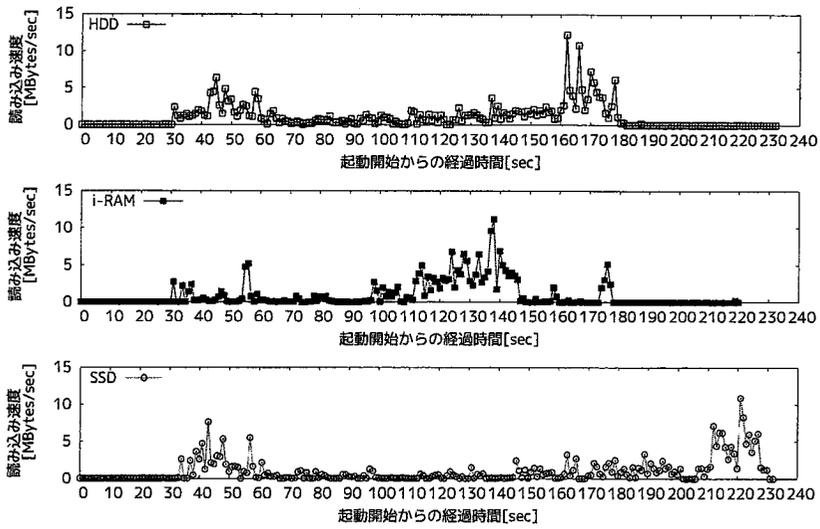


図4 クライアント 22 台同時起動時、ある 1 回のディスク読み込みの計測結果

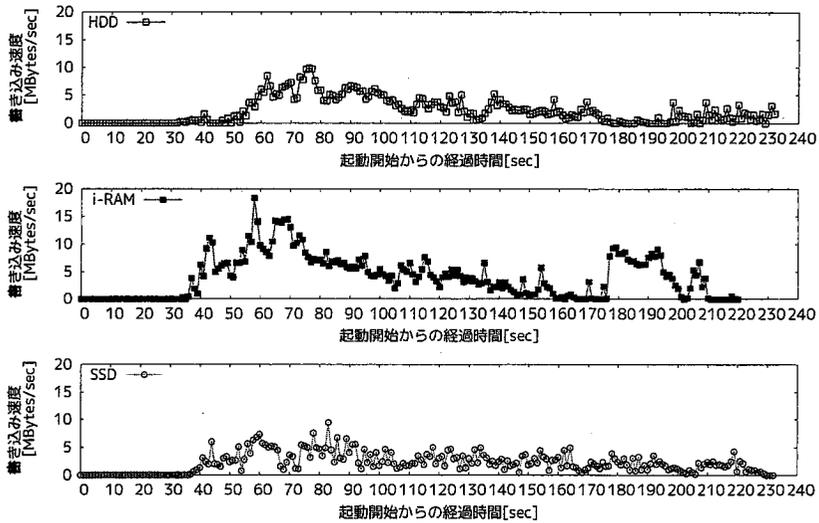


図5 クライアント 22 台同時起動時、ある 1 回のディスク書き込みの計測結果

能差が現れにくいと考えられる。

5.4 ディスク I/O 負荷

図 4 を見る。30 秒から 60 秒にかけてはローダブルモジュールの読み込みを行っていると考えられ、この部分に関してはネットワークと同じく各々のディスクに違いはほとんどないと言える。また、100 秒程度から始まるデーモンの実行に必要なファイルの読み込みが、i-RAM、HDD、SSD の順に遅れているのが見て取れる。この順序は起動時間の順序と同じであり、ディスクの読み込みの遅れが、起動時間に影響を与えているものと考えられる。

図 5 をみると、起動し始めて 60 秒程度経過した頃に、ピークがあり、その後なだらかに書き込みが続いている。i-RAM は 2 度目のピークが 170 から 200 秒程度のところで見られ、210 秒手前のピークを最後に書き込みを終えているのに対し、HDD、SSD は起動完了後も書き込みが続いている。また、この書き込みは後ろにずっと伸びていた。

今回の計測において、ブートサーバのディスクの設定で最終アクセスタイムの更新が行われるようになっている。60 秒前後のあたりではローダブルモジュールを読んでいると考えられるので、この時点でのディスクへの書き込みは、最終アクセスタイムの更新を行っていると考えられる。よって、60 秒以後の書き込みに関しても最終アクセスタイムの更新は行っており、これによる影響があると考えられる。

また、SSD などのフラッシュメモリを用いた製品では書き込みにおいて、ウェアレベリングと呼ばれるフラッシュメモリへの書き込みの平滑化がおこなわれている。⁸⁾ これは書き換え回数の上限があるフラッシュメモリの寿命を延ばすための技術である。今回の計測で SSD が期待される性能が出ていない原因として、読み込んだ一部のファイルに頻繁に書き換えが起り、いわゆるプチフリーズと呼ばれる状態になっているのではないかと考えられる。

この現象は書き込みを行わないようにすることで回避することができる。よって、最終アクセスタイムの更新を行わないようにすることで、性能の低下を緩めることができるのではないかと考える。

6. ま と め

本稿では、ネットワークブートシステムについて、ブートサーバのディスク I/O 性能の向上による性能向上の効果について評価した。クライアント OS のルートファイルシステムを配置するディスクとして、HDD、i-RAM、SSD、RAM ディスクを使用した。ランダムアクセス性能などに優れた RAM や i-RAM を用いることで、クライアントの起動時間が 10~50%削減され、効果があることが実証できた。

また、SSD を用いた場合には、カタログスペックか

ら期待される性能がでなていないことがわかった。

今後の課題として、

- 最終アクセスタイムの更新を行わないように設定した場合にどのようなになるか
- クライアント PC の台数を増やしていった場合どのようなになるか
- NFS のバージョン 4 を用いた場合どのようなになるか
- 計測手法の改善があげられる。

参 考 文 献

- 1) Apple 社:Netboot;<http://www.apple.com/jp/server/macosx/netboot.html>
- 2) Ardence 社:Ardence;<http://www.thinclient-net.com/ardence/about/index.html>
- 3) GIGABYTE : GC-RAMDISK i-RAM ;
http://www.gigabyte.co.jp/Products/Storage/Products_Overview.aspx?ProductID=2180&ProductNave=GC-RAMDISK
- 4) Intel 社 : Pre eXcution Environment ;<http://download.intel.com/design/archives/wfm/downloads/pxespec.pdf>
- 5) Western Digital 社 : WD Caviar Blue Hard Drive Series Specification Sheet ;<http://www.wdc.com/en/library/sata/2879-701277.pdf>
- 6) Super Talent Technology 社 : SSD MasterDrive OX ;http://www.supertalent.com/products/ssd_detail.php?type=MasterDrive%200X#
- 7) 梶田秀夫, 齊藤明紀 : "unionfs を用いたディスクレスシステムの実装とその評価" ; 分散システム/インターネット運用技術シンポジウム 2005 年度論文集 (2005.12)
- 8) Super Talent Technology 社 : SSD IDE Flash Drive White Paper ;
<http://www.supertalent.com/tools/ssd.php>
- 9) GIGABYTE 社 : GC-RAMDISK i-RAM Rev1.3 Manual ;http://asia.giga-byte.com/FileList/Manual/desktop_manual_i-ram_1.3-j.pdf