

Xen 仮想マシン環境における ディスクアクセス性能予測モデルの提案と評価

馬場 輝幸 田中 淳裕*

日本電気株式会社 システムプラットフォーム研究所、*研究企画部

1. はじめに

CPU 等のコンピュータリソースの有効利用やセキュリティの確保のために、仮想マシン (VM: Virtual Machine) 技術の導入が進んでいる。このような VM 環境を運用する際には、割り当てるリソース量と得られる性能の関係を表す性能予測モデルが必要となる。しかし、ディスクに関しては、容量のみが考慮され[1]、アプリケーション性能に影響を与えるアクセススループットの予測については検討されていなかった。

そこで、本稿では、複数 VM で同時にディスクアクセスが生じた時のディスク性能劣化を考慮した性能予測モデルを提案し、実測と比較評価する。なお、本稿では、VM としてオープンソースである Xen を対象とする[2]。

2. 複数ファイル同時アクセス時のディスク性能の問題点

一般的に、複数データにアクセスする際に対象データがディスク上で離れた位置にあると、ヘッドの移動距離であるシーク距離が長くなり、アクセス帯域が低下することが知られている[3]。さらに、複数ファイルへ同時にアクセスが生じた時のシーク距離は OS 等のスケジューラに依存する。例えば、ディスク上で 2 つのファイルのシーク距離が離れており、read 処理を 1 回実行するたびに異なるファイルへ seek する場合と、一方のファイルで read 処理を複数回実行した後に他のファイルへ seek する場合を考えると、前者ではシーク距離が長くなり、スループットは低下する。仮想マシン環境ではリソースの共有度が高くなり、複数データにアクセスする状況が増加するため、上記のような影響を受け易くなることが予想される。しかし、現状では VM 環境の設計時にこのスループットの劣化が考慮できないという問題がある。

そこで本稿では、VM 環境でのディスクスループットを予測するために、複数ファイルへ同時アクセスする時のシーク動作をモデル化した性能予測モデルを提案する。

3. ディスクアクセス性能予測モデル

本章では、シーク距離を考慮したディスク性能予測モデルを説明する。本稿で提案する性能予測モデルは次の特徴を持つ。

- ディスクアクセススケジューラによる動作の違いを表す連続アクセス率を定義。
- 連続アクセス率を重み付け係数として平均シーク距離を算出。
- 上記で計算した平均シーク距離を元にディスク固有の性能情報 (シーク距離と帯域の関係) からアクセス帯域を算出。

以下で、図 1 を参照して、各特長の詳細を説明する。はじめに、連続アクセス率を定義する。連続アクセス率は、複数ファイルアクセス時に同一ファイルを連続してアクセスする確率を表し、0 から 1 の範囲の値をとる。連続アクセス率=0 の場合は、read 処理を 1 回実行した後は必ず他のファイルへシークが発生するため、スループットは低下する。逆に、連続アクセス率=1 の場合は、同一ファイルのみを連続してアクセスすることとなり、シーク距離が短くなりスループットは向上する。一方で、ある限られた時間範囲内の各ファイルのスループットに注目すると、連続アクセス率が高い場合は同一ファイルに連続してアクセスするためファイルごとのスループットにバラつきが生じ、反対に連続アクセス率が低い場合は各ファイルのスループットが均等になる。

上記のように連続アクセス率を定義し、他ファイル (全ファイル数を N とすると、自ファイルを除いた $N-1$ 個) へのアクセス確率が等しいとすると、他の各ファイルへのアクセス確率は $(1-\alpha)/(N-1)$ となる。この連続アクセス率は、Linux や Xen 等のプラットフォームによって値が定まる。

続いて、連続アクセス率を重み係数とした平均シーク距離の算出方法を説明する。

ファイル i とファイル j の間のファイル間の平均シーク距離を l_{ij} 、ファイル i 内での平均シーク距離をファイル内平均シーク距離 l_{ii} で表す (図 1)。ランダムアクセスの場合は、ファイル間平均シーク距離はファイル i の中心位置とファイル j の中心位置の間のシーク距離、ファイル内平均シーク距離は、ファイルサイズの $1/2$ となる。

以上の条件において、複数ファイルへランダムアクセスする際の平均シーク距離 l_{ave} は式(1)で求まる。

$$l_{ave} = \frac{1}{N} \left(\alpha \sum_{i=1}^N l_{ii} + \frac{1-\alpha}{N-1} \sum_{i=1}^N \sum_{j \neq i} l_{ij} \right) \quad \text{式(1)}$$

最後に、この平均シーク距離からアクセス帯域を求める。個別ディスクの性能情報として予め測定しておくシーク距離 (l) とアクセス帯域 (B) の関係式 $B=f(l)$ から、上記で算出した平均シーク距離 (l_{ave}) を l とした際のアクセス帯域 B の値を予測帯域 B_{est} として求める。

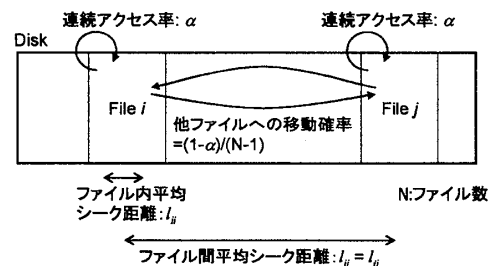


図 1: 連続アクセス率と平均シーク距離の説明図

4. 性能予測モデル評価

本章では、提案する性能予測モデルを評価するために、Xen による複数 VM 環境でディスクスループットを実測し、モデルによる予測値と比較する。また、仮想マシン環境でない通常の Linux(以後 native Linux と記述)でも測定し、Xen 環境と比較する。

図 2 に測定系を示す。マシンスペックの影響を避けるために、1 台の物理マシンに、native Linux と、Xen 環境をデュアルブートで構築した。物理マシン及び native Linux、Xen のスペックは表 1 に示した通りである。また、一般的にディスク上でのファイル位置は不明であるが、本稿ではファイルを希望の位置に固定するために、10GB ごとにパーティションで分割し、パーティション内に Xen VM イメージファイルを作成した。read 対象ファイルは、VM 内で dd コマンドを利用して乱数を書き込み、2GB のファイルを作成した($l_i=1GB$)。なお、シーク距離は本来シリンダ数で表されるが、本稿ではユーザにとって扱い易い Byte 単位で表す。

Native Linux 及び Xen 仮想マシン環境において、ランダム read を実行して、そのアクセス帯域を測定した。ランダム read は、2GB の対象ファイル内を一様ランダムに seek して、一定サイズ(4kB)を read するという動作を繰り返すプログラムである。Native Linux 環境では、ランダム read を複数プロセスとして実行し、Xen VM 環境では、複数 VM を構築し、各 VM で 1 つのランダム read プロセスを実行した。

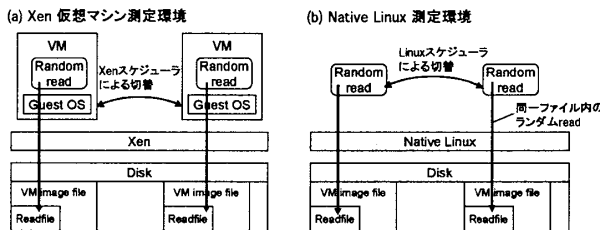


図 2: 実験構成

表 1: 実測に用いた物理マシン及び OS のスペック

	Native Linux	Xen (version 3.0.4)	
		Dom 0	Dom U
CPU	Pentium 4 3.8GHz		
Disk	160 GB (Western Digital, WD1600JS)		
Memory	3 GB	1 GB	256 MB
OS (Kernel)	Fedora Core 6 (2.6.18)	Fedora Core 6 (2.6.18)	Fedora Core 6 (2.6.18)

測定結果の一例として、ランダム read プロセス数=2、read 対象ファイル間のシーク距離を 20 GB ($l_{ij}=20GB$) に固定したときの測定結果を図 3 に示す。スループットが安定した時間領域での 2 プロセスのトータルアクセス帯域は Xen では 0.264 MB/sec、native Linux では 0.389 MB/sec となった。Xen の場合、スループットは小さいが、ファイル間でのスループットのバラつきが小さい結果となり、Xen での連続アクセス率が低いことを示している。

図 4 に、平均シーク距離(l_{ave})の変化に対するスループットの予測値と実測値を示す。ランダム read プロセス数を 1、2、4、8 とし、横軸を平均シーク距離、縦軸をトータルアクセス帯域としている。異なる平均シーク距離(l_{ave})で測定するために、パーティションを利用して read 対象ファイルの位置(式(1)中の l_{ij})を変えて測定した。平均シーク距離を求めるための連続アクセス率は、2 プロセス時の測定結果を元にしたフィッティングから、native

Linux と Xen で、それぞれ 0.99、0.0 とした。また、今回利用したディスクの性能としてシーク距離と帯域の関係は、ファイル内平均シーク距離($l_{ij}=1GB$)に対しては $f(l_{ij})=0.453$ MB/sec、ファイル間平均シーク距離($l_{ij}>20GB$)に対しては $f(l_{ij})=1/(0.0244 \times l_{ij} + 2.91)$ であった。

図 4 の結果から、native Linux 環境及び、Xen 環境のそれぞれにおいて、提案する予測方式が誤差 10% でトータル帯域を予測できることが確認できた。

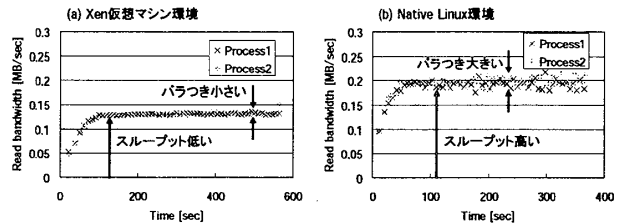
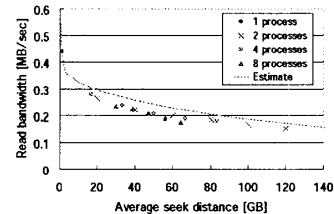


図 3: プロセス数=2 の場合の帯域測定結果 ($l_{ij}=20GB$ 固定)

(a) Xen 仮想マシン環境 (連続アクセス率=0.0)



(b) Native Linux 環境 (連続アクセス率=0.99)

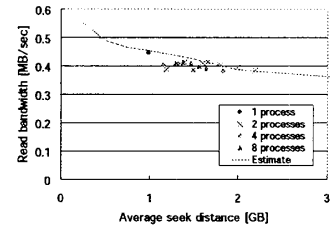


図 4: 平均シーク距離ごとの予測値と実測値の比較

5. まとめ

本稿では、複数ファイルを同時にアクセスする際のアクセス帯域の予測方法を提案した。本方式では、同一ファイルを連続してアクセスする確率を連続アクセス率として定義し、この値を利用することで異なる OS 環境であっても統一的に予測することが可能となる。また、実測により本予測方法が誤差 10% で予測可能なことを示した。本稿ではローカルディスクを対象としたが、今後は複数ディスクで構成される SAN (Storage Area Network) を対象にする予定である。

謝辞

本研究の一部は、総務省の委託研究「次世代バックボーンに関する研究開発」プロジェクトの成果である。

参考文献

- [1] Y. Ding and E. Bolker, "How Many Guests Can You Serve? - On the Number of Partitions," in *Proceedings of CMG conference*, 2006.
- [2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the Art of Virtualization," in *Proceedings of the 19th ACM SOSOP*, pp. 164 - 177, October 2003.
- [3] C. Ruemmler and J. Wilkes, "An Introduction to Disk Drive Modeling", *Computer*, Vol. 27, No. 3, pp. 17-29, 1994.