

様々な仮想計算機におけるディスク入出力性能の比較

佐藤 将也^{1,a)} 谷口 秀夫¹ 鷗島 匠²

概要：仮想計算機では，ディスク入出力において仮想計算機モニタを介して処理が行われる．このため，入出力処理の性能は，仮想化による影響を受ける．また，仮想計算機の構成法は仮想計算機モニタにより異なる．仮想計算機の構成が異なると，入出力処理を行う際に利用されるプログラムやモード遷移が異なるため，構成により入出力処理の性能に影響がある．さらに，仮想計算機の構成が同じ場合でも，入出力処理に用いられるプログラムが異なることで性能への影響がある．本稿では，物理計算機と様々な仮想計算機におけるディスク入出力処理の性能を測定し，仮想計算機の構成法や利用されるプログラムによる処理性能への影響を評価した結果を報告する．

1. はじめに

仮想計算機 (Virtual Machine, 以降 VM) の普及にともない，様々なプログラムが VM で実行されている．VM は様々な構成法があり，仮想計算機モニタ (Virtual Machine Monitor, 以降 VMM) によって性能も異なる．例えば，Xen [1] のようなベアメタル型の VMM では，すべてのオペレーティングシステム (Operating System, 以降 OS) が VM で動作し，入出力は VMM と特権 VM を介して行われる．また，VMware Workstation [2] のようなホスト型の VMM では，VMM がホスト OS の応用プログラム (Application Program, 以降 AP) として動作し，VM 上の OS (以降，ゲスト OS) は，VMM とホスト OS を介して入出力を行う．さらに，KVM [3] や bhyve [4] のように，Intel VT [5] や AMD-V [6] などの CPU に搭載された仮想化支援機能を前提として，入出力処理をホスト OS の AP で行う構成の VMM も利用されている．このように，VMM の構成により入出力処理の形態が異なる．仮想化環境における入出力性能はこれまでに評価されている [7], [8], [9], [10] もの，応用プログラムの処理性能に着目したものが多く，基礎的な性能における比較評価は充分ではない．特に，bhyve は他の VMM と比較して報告は少ない．

本稿では，物理計算機 (Bare-metal Machine, 以降 BM) と VM におけるディスク入出力処理の性能の違いを評価した結果を述べる．また，VM の評価では，様々な VMM においてディスク入出力の処理時間を評価し，VMM の構成の違いによる性能への影響を評価した結果を述べる．

2. 仮想計算機の構成

評価では，VMM として Xen，VMware Workstation，KVM，および bhyve を用いる．図 1 に評価に用いる VMM を示し，以下でそれぞれの特長を述べる．

(1) Xen [1]

Xen は，ベアメタル型の VMM であり，すべての OS が VM で動作する．管理用の VM が 1 つあり，他の VM からの入出力は VMM を介して管理 VM の AP が行う．Xen は，準仮想化と完全仮想化の 2 方式を提供する．完全仮想化環境では，管理 VM で動作する AP (QEMU) を介して入出力を行う．また，完全仮想化環境において，管理 VM と VM 間で共有メモリを介して準仮想化方式による入出力を行う方法もある．

(2) VMware Workstation [2]

VMware Workstation は，ホスト型の VMM であり，AP としてホスト OS で動作する．ここでは，ホスト OS として Linux を想定する．VM からの入出力は VMM とホスト OS を介して行われる．

(3) KVM [3]

KVM は，ホスト OS で動作する VMM であり，Linux のカーネルモジュールと入出力を行う AP (QEMU) の組み合わせにより実現されている．VMware Workstation と同様にホスト型 VMM と考えられるが，Intel VT などの CPU による仮想化支援機能を利用してゲスト OS をゲストモードで動作させる点が異なる．

(4) bhyve [4]

bhyve は，ホスト OS で動作する VMM であり，FreeBSD のカーネルモジュール (vmm.ko) と入出

¹ 岡山大学 大学院自然科学研究科

² 岡山大学 工学部

a) sato@cs.okayama-u.ac.jp

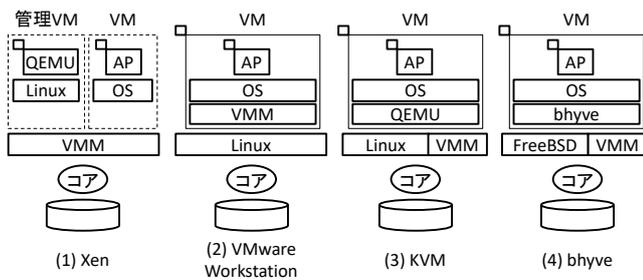


図 1 評価対象の VMM

力を行う bhyve の組み合わせにより実現されている。基本的な構造は KVM と同様であり、Intel VT などの CPU による仮想化支援機能を利用する。

KVM と bhyve は同種の VMM だが、入出力を行う AP が異なるため、両者の処理性能を比較することにより、実装の違いが入出力処理の性能に与える影響を評価する。

3. 評価

3.1 観点

評価の観点を以下に示す。

(観点 1) 仮想化の有無による処理性能の違い

仮想化の有無による入出力の処理性能への影響を評価する。VMM は VM に仮想的なハードウェアを提供し、VM の動作を管理する。BM と VM では、VMM の介在の有無により入出力処理の性能に違いが出る。そこで、仮想化の有無による処理性能の違いを評価する。

(観点 2) VMM の構造による処理性能の違い

VMM は 2 章で述べたように様々な構成がある。構成により、入出力処理に介在するプログラムや処理が異なるため、VMM の構造による処理性能の違いを評価する。

(観点 3) シングルコアとマルチコアによる処理性能の違い

VMM は、VM に対して CPU を 1 つまたは 2 つ以上を割り当てることができる。このとき、割り当てるコア数によって処理性能に違いがあると考えられる。このため、シングルコアとマルチコアにおける処理性能の違いを評価する。

(観点 4) 走行させる VM 数による処理性能の違い

仮想化環境では、1 台の計算機で複数の VM を実行できる。VM による入出力には VMM が介在するため、複数の VM が走行することで処理性能に影響すると考えられる。このため、走行させる VM 数による処理性能の違いを評価する。

3.2 測定の環境と内容

評価環境を表 1 に示す。評価では、CPU は Intel Xeon E5-2609 v4、メモリは 32 GB、ディスク（以降 DK）は

表 1 評価環境

H/W	CPU	Intel Xeon E5-2609 v4 (1.7 GHz, 8-cores)
	メモリ	32 GB
	DK	Western Digital WD1005FBYZ (SATA600, 7,200rpm, 128 MB Cache)
BM	OS	FreeBSD 11.0-RELEASE (64bit)
VM	VMM	Xen, VMware Workstation, KVM, bhyve
	ホスト OS	Xen: 無し VMware Workstation: Debian 9.0.0 (Linux 4.9.0, 64bit) KVM: Debian 9.0.0 (Linux 4.9.0, 64bit) bhyve: FreeBSD 11.0-RELEASE (64bit)
	ゲスト OS	FreeBSD 11.0-RELEASE (64bit) (Xen の管理 VM: Debian 9.0.0 (Linux 4.9.0, 64bit))
	仮想 CPU	1 or 2
	メモリ	1 GB
	仮想 DK	16 GB (単一ファイル)

Western Digital WD1005FBYZ を搭載した計算機を用いた。仮想 CPU は、割当数による性能への影響を評価するために、1 つまたは 2 つを割り当てた。なお、仮想 CPU は、実 CPU に対して固定せず動的に割り当てる。ゲスト OS は FreeBSD 11.0-RELEASE を用いた。ホスト OS は Debian 9.0.0 を用い、bhyve の評価のみ FreeBSD 11.0-RELEASE を用いた。これは、bhyve が FreeBSD を対象として開発されているためである。なお、Xen における管理 VM の OS、および VMware Workstation と KVM におけるホスト OS は Debian 9.0.0 とし、マルチユーザーモード（Linux の場合はマルチウィンドウモード）で動作させた。ゲスト OS には 1 GB のメモリを割り当て、仮想 DK はホスト OS の単一の通常ファイルとして 16 GB を割り当てた。

評価では、BM または各種 VMM (Xen, VMware Workstation, KVM, および bhyve) で動作する VM について、コア数 (1 コアまたは 2 コア) と VM 数 (1VM または 2VM) による入出力性能を比較する。

評価で用いる入出力処理は、RAW デバイスに対する読み込み処理 (ランダムリード) と書き込み処理 (ランダムライト) とする。入出力サイズは 4, 16, および 32 KB とする。なお、測定プログラムのみが動作する環境では、測定プログラムが計算機資源を占有してしまい、仮想化による影響の評価が難しいことから、測定プログラムに加えて負荷プロセスを走行させた。負荷プロセスは、測定プログラムと同様の処理を常に繰り返す。負荷プロセス数は、0, 2, および 4 とする。VM 数を 2 としたときの評価では、同時に走行する 2VM のうち、片方の VM でのみ評価を行う。

測定では、測定プログラムが読み込み処理または書き込み処理を 100 回繰り返した際の 1 回ごとの処理時間を用

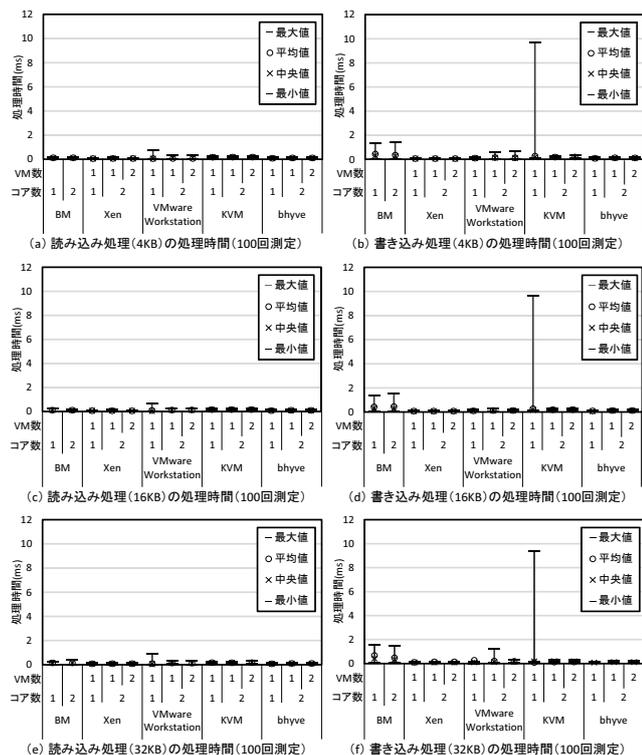


図 2 BM と VM におけるディスク入出力性能の比較 (負荷プロセス数 0)

いた。また、読み込み処理または書き込み処理でアクセスする範囲は、事前に 100 個の乱数を生成しておき、RAW デバイスの先頭アドレスから 0-2 GB のオフセットを指定した。

3.3 結果と考察

負荷プロセス数を 0, 2, および 4 とした場合の BM および VM のディスク入出力性能の測定結果をそれぞれ図 2, 図 3, および図 4 に示す。

評価結果より、以下が分かる。

(1) 仮想化の有無

(a) BM における処理時間の最大値は、共存プロセス数が 2 または 4 のとき、各種 VM における処理時間の最大値と比較して 2 倍大きい。一方、共存プロセス数が 0 のときは、この傾向は見られなかった。この原因を調査するために、BM かつ 1 コアの環境で、読み込み処理 (4 KB) を行うときの 1 回ごとの処理時間を分析した。図 5 に測定プログラムと負荷プロセス (1, 2, および 4) を走行させた際の読み込み処理時間を示す。この結果より、一定の周期で処理時間が大きくなっており、負荷プロセス数が大きくなるほど、1-50 回目の範囲で処理時間の大きくなる回数が増えることが分かる。また、60-90 回目あたりでは、負荷プロセス数に関わらず処理時間が小さくなっていることが分かる。これは、負荷プロセスが OS 処理を行っている場合、測定プロセスが READY 状態となりすぐには実行されないためだと考えられる。VM では、後述するように、ホ

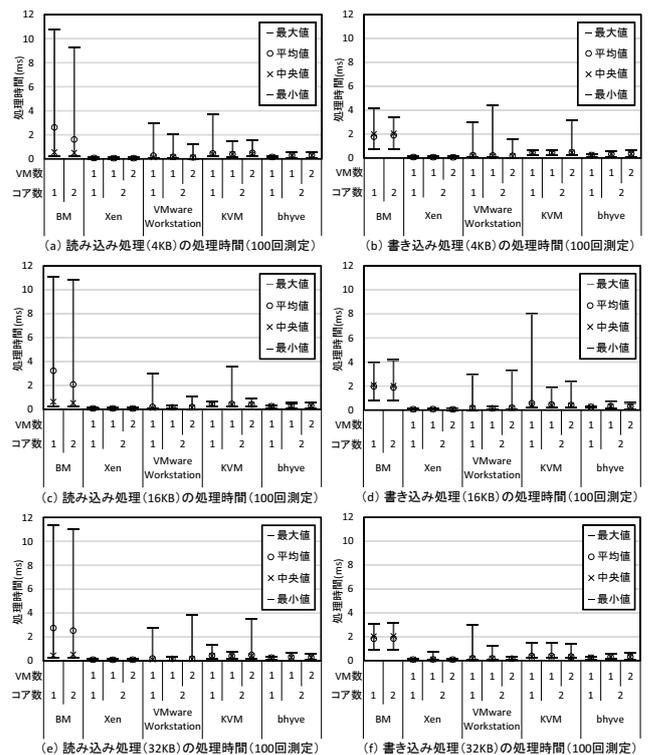


図 3 BM と VM におけるディスク入出力性能の比較 (負荷プロセス数 2)

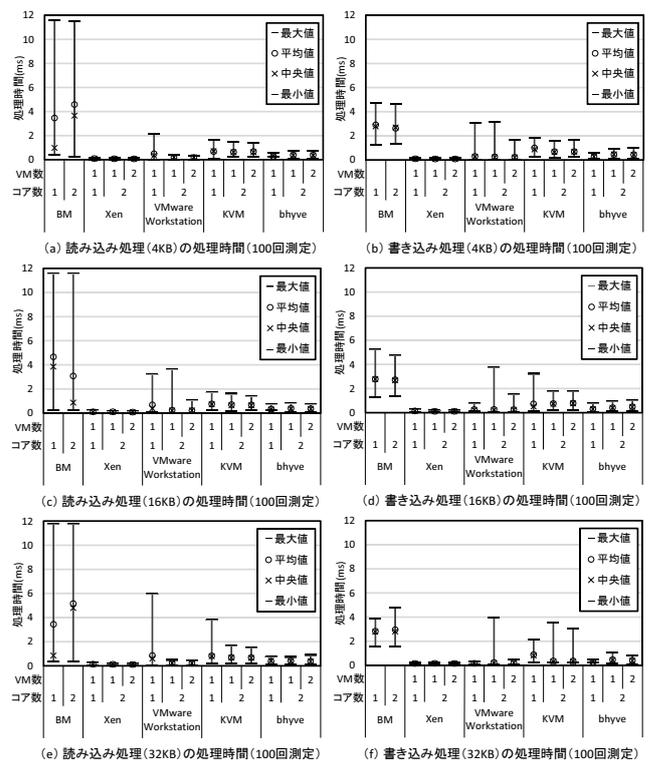


図 4 BM と VM におけるディスク入出力性能の比較 (負荷プロセス数 4)

スト OS や管理 VM の OS が持つバッファキャッシュにより入出力の処理時間が短くなるため、この現象が低減されていると考えられる。

(b) 図 4 より、VM は BM より読み込み処理時間が短い傾

向にある．また，VMの書き込み処理時間の最小値はBMより小さい．これは，BMの入出力はRAWデバイスアクセスであるのに対して，VMの入出力はホストOSまたは管理VMのOSにおける通常ファイルへのアクセスとなるためである．図6にBM環境におけるRAWデバイスと通常ファイルに対する入出力処理の時間を示す．図6のように，RAWデバイスへの読み書きに対して通常ファイルへの読み書きの処理時間は大幅に短いため，VMにおけるRAWデバイスへの読み書きは，ホストOSや管理VMのOSが持つバッファキャッシュにより短くなる．

(2) VMMの構造

(a) 図2, 図3, および図4より, Xenは他のVMMより処理時間が短く, 最大値と最小値の差が小さい．特に, 図3と図4においてこの傾向が強く見られる．これは, FreeBSDがXenにおける非同期入出力に対応していたためだと考えられる．FreeBSD 11.0-RELEASEでは, Xenにおける準仮想化に対応した入出力を行うドライバを搭載しており [11], この影響でXenにおける処理時間が短くなったと考えられる．準仮想化に対応した入出力では, 特権VMとVMで共有メモリを利用して非同期に入出力を行うため, 他VMMよりも処理時間が短くなったと推測できる．

(b) 図2, 図3, および図4より, KVMの測定結果は, bhyveと比較して最大値と最小値の差が大きい．このことから, VMMの構造が同じであっても, 実装による入出力処理の性能への影響があることが分かる．

(c) 図3より, KVM/1コア/1VMの書き込み処理時間の最大値は10ms程度である．これは, 他の測定結果より極端に大きい．この原因を調査するために, KVM/1コア/1VMの場合について, ホストOSをシングルユーザモード, マルチユーザモード, およびマルチウィンドウモードで動作させた場合の処理時間を測定した．図7に測定結果を示す．測定結果より, シングルユーザモードとマルチユーザモードでは大きな差はない．しかし, ホストOSをマルチウィンドウモードで動作させた場合に, 処理時間の最大値が極端に大きくなっていることが分かる．これは, マルチウィンドウモードでは, ホストOSにおいてVMM以外のプロセスが多く動作しているためだと考えられる．

(3) シングルコアとマルチコア

(a) 図2における処理時間0-2msの部分の拡大図を図8に示す．図8より, Xenとbhyveについて, 1コア/1VMと2コア/1VMの結果を比較すると, 1コア/1VMの方が処理時間の平均値が小さい．一方, VMware WorkstationとKVMではこの傾向は見られなかった．このことから, 一部のVMMにおいて, 負荷が大きくない処理を行うときは, シングルコアの方がマルチコアよりも処理性能が高くなる場合がある．

(4) 走行させるVM数

(a) 図8より, 各種VMMについて, 2コア/1VMと2コ

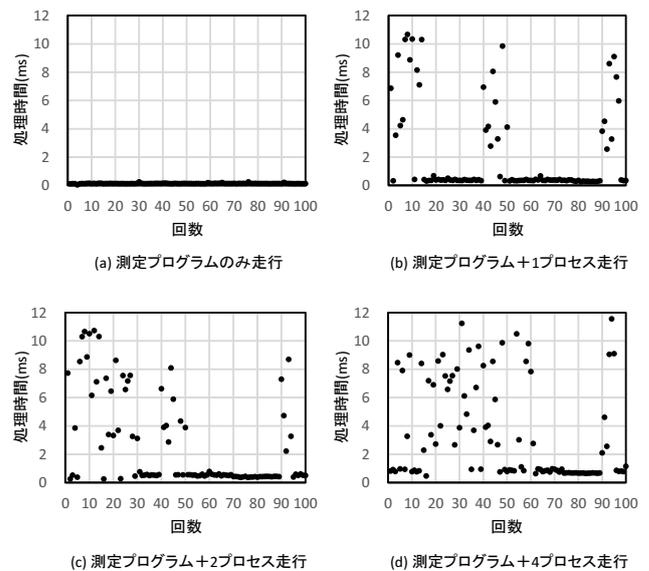


図5 BM/1コアにおける負荷プロセス数ごとの読み込み処理(4KB)時間

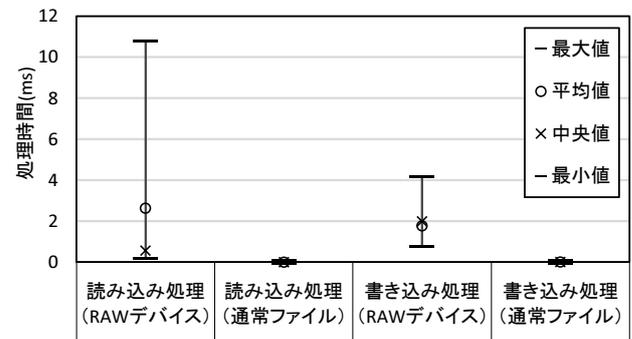


図6 BM/1コアにおけるRAWデバイスと通常ファイルへの入出力処理時間(入出力サイズ4KB, 負荷プロセス数2)

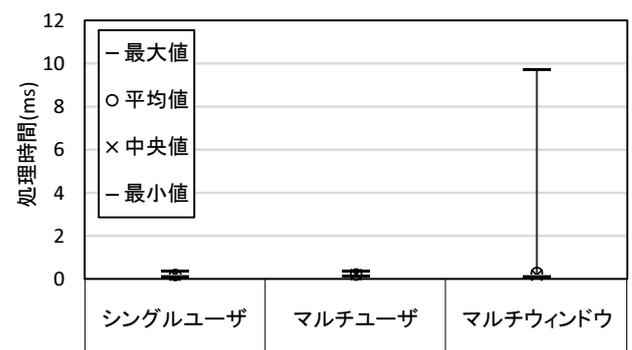


図7 KVMにおけるホストOSの動作モードごとの1コア/1VMの場合の書き込み処理時間

ア/2VMの結果を比較すると, 測定結果に大きな差はない．このことから, 測定対象以外のVMで負荷が大きくない場合は, VM数を2つに増やしても入出力性能への影響は殆ど無いことが分かる．

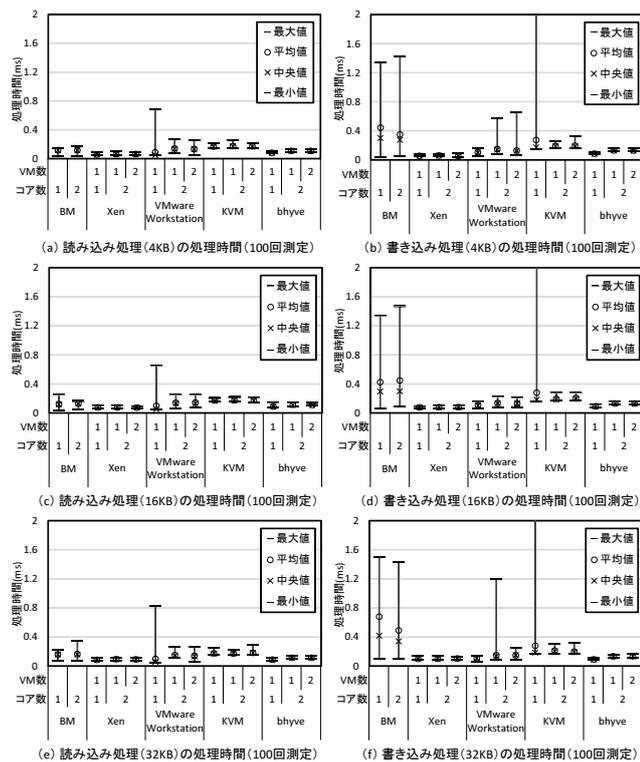


図 8 BM と VM におけるディスク入出力性能の比較 (負荷プロセ
ス数 0, 処理時間 0~2ms の範囲を拡大)

4. 関連研究

Hwang ら [7] は, 4 種類の VMM を用いて性能評価を行っている. ファイル入出力の評価では, KVM が Xen より高速な結果となっている点が本報告の結果と異なる. これは, 準仮想化ドライバの有無によるものだと考えられる. また, ホスト型 VMM や bhyve については評価されていない. Pu ら [8] は, Xen において複数 VM が通信を行ったときに, ある VM の処理が他 VM の処理性能に与える影響が評価され, CPU 負荷とネットワーク負荷によって性能低下の性質が異なることが示されている. 本稿の評価プログラムはディスク入出力を繰り返すものであり, CPU 負荷を考慮した評価が必要である. また, 複数 VM を利用した評価では, 他 VM において, 評価結果に影響するような CPU 処理や入出力処理を行っていない. このため, 他 VM において CPU 処理や入出力処理を行った際の影響を評価する必要がある.

VM における I/O スケジューラの有効性が評価されている [9]. 評価により, ゲスト OS における I/O スケジューリングによって, モード遷移の回数が削減され, 仮想化によるオーバーヘッドの抑制に繋がることが指摘されている. また, VMM 層とゲスト OS それぞれが協調して I/O スケジューリングを行うことで, 仮想化環境に適した性能改善の可能性が示されている. 組み込み環境では, 資源を VM ごとに隔離できることから, VMM の利用が研究されてい

る. Sañudo ら [10] は, VM の I/O スケジューリングについて, 準仮想化ドライバを用いることでスケジューリングが困難になる可能性を指摘している.

Xen や KVM では準仮想化入出力として PVOPS [12] や virtio [13] が提案されている. 本報告における評価では, Xen を用いた場合に PVOPS の効果が大きかった. virtio は KVM や bhyve など様々な VM で利用可能であるが評価が充分でない. このため, 残された課題として, virtio を利用した環境における評価がある.

VM とコンテナにおける性能評価が行われている [14]. 評価では, VMM として KVM と virtio を用いて VM からブロックデバイスにアクセスした際の性能が BM や Docker と比較されている. ランダムリードとランダムライトにおけるスループットの評価結果では, BM と Docker はほぼ同じ性能だが, KVM は約半分の性能だと報告されている. また, I/O 命令の実行時に, BM や Docker と比較してより多く CPU を使用するため, AP 処理に CPU を使用できないことが指摘されている. Morabito ら [15] は, コンテナや従来の VMM を用いた仮想化環境だけでなく, Library OS である OSv [16] を用いた性能評価を行っている. Library OS は, VM 上で動作することを前提に作成された OS であり, VM による隔離を維持しつつ OS 処理の再現による性能低下を抑制している. Morabito らの評価では, 入出力処理の性能評価に Bonnie++ を用いているが, OSv は Bonnie++ で用いられる fork() システムコールに対応していないため, 測定が行われていない.

5. おわりに

物理計算機と仮想計算機における入出力の処理性能を比較評価した. 評価では, 物理計算機と様々な VMM において, ディスクに対するランダム読み込みとランダム書き込みの処理時間を比較評価した.

評価結果より, 仮想計算機では, ホスト OS または管理 VM の OS におけるバッファキャッシュの影響により, 処理時間が短い傾向にあった. また, VMM の構造により, 入出力の処理時間が大きく異なることが分かった. 特に VMM として Xen を用いた場合, 準仮想化ドライバがゲスト OS の FreeBSD に搭載されており, 入出力処理が特権 VM の OS と共有メモリを介して非同期で行われるため, 入出力処理が高速となった. また, KVM は, ホスト OS をマルチウィンドウモードで動作させたときに, 処理時間の最大値が極端に大きくなる場合があることが分かった. さらに, シングルコアとマルチコアでは, Xen と bhyve においては処理負荷が大きくない場合はシングルコアの方が処理性能が高くなる場合があった. 走行させる VM 数による処理性能は, 測定対象の VM 以外の負荷が大きくない場合は影響が殆ど無いことを示した.

残された課題として, 通信処理や応用プログラムを用い

た評価がある。

謝辞 本研究の一部は JSPS 科研費 18K18051 の助成を受けたものです。

参考文献

- [1] Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neugebauer, R., Pratt, I. and Warfield, A.: Xen and the Art of Virtualization, *SIGOPS Oper. Syst. Rev.*, Vol. 37, No. 5, pp. 164–177 (online), DOI: 10.1145/1165389.945462 (2003).
- [2] VMware: Workstation Pro, (online), available from <https://www.vmware.com/products/workstation-pro.html> (accessed 2020-1-30).
- [3] KVM: Main Page — KVM, (online), available from https://www.linux-kvm.org/index.php?title=Main_Page&oldid=173792 (accessed 2020-1-30).
- [4] Dexter, M.: BSD Hypervisor, (online), available from <https://bhyve.org/> (accessed 2020-1-30).
- [5] Intel: Intel Virtualization Technology (Intel VT), Intel (online), available from <https://www.intel.com/content/www/us/en/virtualization/virtualization-technology/intel-virtualization-technology.html> (accessed 2020-2-3).
- [6] AMD: AMD-V Technology for Client Virtualization, AMD (online), available from <https://www.amd.com/en/technologies/virtualization> (accessed 2020-2-3).
- [7] Hwang, J., Zeng, S., y Wu, F. and Wood, T.: A Component-Based Performance Comparison of Four Hypervisors, *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, IEEE, pp. 269–276 (2013).
- [8] Pu, X., Liu, L., Mei, Y., Sivathanu, S., Koh, Y. and Pu, C.: Understanding Performance Interference of I/O Workload in Virtualized Cloud Environments, *2010 IEEE 3rd International Conference on Cloud Computing*, IEEE, pp. 51–58 (2010).
- [9] Boutcher, D. and Chandra, A.: Does virtualization make disk scheduling passé?, *ACM SIGOPS Operating Systems Review*, Vol. 44, No. 1, pp. 20–24 (2010).
- [10] Sañudo, I., Cavicchioli, R., Capodieci, N., Valente, P. and Bertogna, M.: A survey on shared disk I/O management in virtualized environments under real time constraints, *ACM SIGBED Review*, Vol. 15, No. 1, pp. 57–63 (2018).
- [11] The FreeBSD Documentation Project: FreeBSD 11.0-RELEASE Release Notes, The FreeBSD Project (online), available from <https://www.freebsd.org/releases/11.0R/relnotes.html> (accessed 2020-2-3).
- [12] Xen Project: XenParavirtOps, (online), available from <https://wiki.xenproject.org/wiki/XenParavirtOps> (accessed 2020-2-3).
- [13] Russell, R.: virtio: Towards a De-Facto Standard for Virtual I/O Devices, *ACM SIGOPS Operating Systems Review*, Vol. 42, No. 5, pp. 95–103 (2008).
- [14] Felter, W., Ferreira, A., Rajamony, R. and Rubio, J.: An Updated Performance Comparison of Virtual Machines and Linux Containers, *2015 IEEE international symposium on performance analysis of systems and software (ISPASS)*, IEEE, pp. 171–172 (2015).
- [15] Morabito, R., Kjällman, J. and Komu, M.: Hypervisors

- vs. Lightweight Virtualization: a Performance Comparison, *2015 IEEE International Conference on Cloud Engineering*, IEEE, pp. 386–393 (2015).
- [16] Kivity, A., Laor, D., Costa, G., Enberg, P., Har’El, N., Marti, D. and Zolotarov, V.: OSv—Optimizing the Operating System for Virtual Machines, *2014 USENIX Annual Technical Conference (USENIX ATC 14)*, Philadelphia, PA, USENIX Association, pp. 61–72 (online), available from <https://www.usenix.org/conference/atc14/technical-sessions/presentation/kivity> (2014).