

# 物理マシン間のライブマイグレーション手法の提案

深井 貴明<sup>†</sup> 表 祐志<sup>†</sup>  
品川 高廣<sup>††</sup> 加藤 和彦<sup>†</sup>

## 1. はじめに

OS のライブマイグレーション<sup>1)</sup>とは、ある物理マシンで動作している OS を停止することなく別の物理マシン上へ移動する機能である。クラウドコンピューティングにおいて、OS のライブマイグレーションはメンテナンスや負荷分散のために用いられる重要な機能である。例えば、物理マシンのメンテナンスをおこなう際、メンテナンス対象の物理マシン上で動作するサービスを別のマシンに移す必要がある。この時に、OS をライブマイグレーションによって他の物理マシンへ移動することで、サービスを停止することなく物理マシンをメンテナンスできる。

現在、OS のライブマイグレーションは仮想化技術によって実現されている。従来の仮想化技術ではハードウェアを完全に仮想化するため、一定のオーバヘッドが生じる。このオーバヘッドは、データベースや HPC などの高負荷な計算の用途には無視できないものとなる<sup>3)</sup>。そこで、デバイスを仮想化せず、OS に物理デバイスを直接アクセスさせることによって、仮想化によるオーバヘッドを削減する手法<sup>2)</sup>が提案されている。

OS が直接デバイスへアクセスすることにより、性能は改善するものの、ライブマイグレーションは困難になる。なぜなら、OS が物理デバイスへ直接アクセスすると、仮想マシンモニタは OS がアクセスしているデバイスの状態を把握できないためである。

既存研究における物理マシン環境上の OS のライブマイグレーション手法は、OS やファームウェアに依存する手法である<sup>4)5)6)</sup>。OS に依存する手法は、手法を実現するための変更をなされた OS でしか利用でき

ない。このため、クラウドの顧客が利用可能な OS を制限する。同様に、ファームウェアに依存する手法は、手法を実現するための変更をなされたファームウェアが存在するハードウェアでしか利用できない。このため、クラウド事業者が利用可能なハードウェアを制限する。

そこで、我々は、OS やファームウェアに依存せずに、物理マシン上の OS をライブマイグレーションする手法を提案する。提案手法では、準パススルー型仮想マシンモニタを用いる。準パススルー型仮想マシンモニタとは、デバイスを仮想化せず、OS からハードウェアへのアクセスを可能な限りパススルーしつつ、一部のアクセスを捕捉する仮想マシンモニタである。この準パススルー型仮想マシンモニタを用いて、移動元マシンの CPU、メモリ、物理デバイスの状態を取得し、移動先のマシンでこれらの状態を復元する。仮想マシンモニタを用いることで、OS やファームウェアに依存せずに、OS のライブマイグレーションを実現する。また、準パススルー型仮想マシンモニタはデバイスを仮想化しないため、仮想化のオーバヘッドも削減できる。

## 2. デバイス状態のマイグレーション

物理デバイスの状態を取得する方法としてまず考えられるのは、デバイスのレジスタから値を読み出す方法である。しかし、デバイスは一部の状態を書き込み専用のレジスタに保持するため、レジスタから値を読むだけでは取得できない状態がある。

本手法では、書き込み専用レジスタが保持するデバイスの情報を取得するため、OS が発行する I/O を監視、記録する。この方法によるデバイス状態のマイグレーションについて、図 1 に示し、以下で説明する。

書き込み専用レジスタへの書き込みの監視は、OS 起動開始時からおこなう。通常時、準パススルー型仮想マシンモニタは、OS からデバイスへ発行される I/O のうち、書き込み専用レジスタへの書き込み I/O のみを監視、記録する。マイグレーション時には、レジ

<sup>†</sup> 筑波大学大学院システム情報工学研究科コンピュータサイエンス専攻

Department of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba

<sup>††</sup> 東京大学情報基盤センター

Information Technology Center, The University of Tokyo

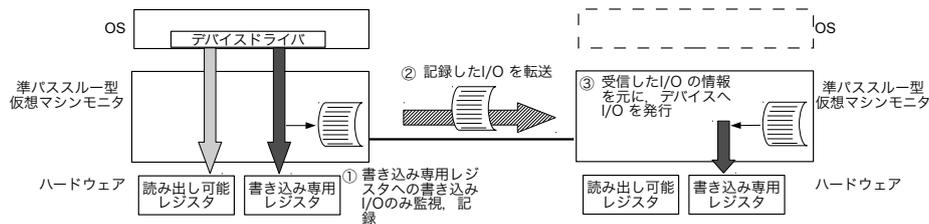


図 1 デバイス状態のマイグレーション

スタの情報に加え、記録した I/O 命令の情報も転送する。移動先マシンでは、レジスタ情報を書き込みに加え、移動元で記録した I/O を実際に発行し、デバイスの状態を復元する。

デバイスが書き換える状態については、I/O を監視することでは得られない。しかし、デバイスが書き換える状態のうち OS の動作に影響を与えるような状態は、OS が把握する必要があるため、読み出し可能なレジスタへ格納されると考えられる。このため、デバイスが書き換える状態の取得はレジスタから値を読み出すことで得られるもので十分である。

### 3. 実装状況

現在、準バスルー型仮想マシンモニタである BitVisor<sup>7)</sup> を用い、実際に割り込みコントローラ (PIC) やタイマーデバイス (PIT) について状態を取得・復元する機構のプロトタイプを実装し、機能を制限した Linux をマイグレーションできることを確認している。

### 4. まとめ

本研究では、物理マシン上での OS ライブマイグレーションを実現する手法を提案した。準バスルー型仮想マシンモニタを用い、OS やファームウェアに依存せずにライブマイグレーションを実現する手法とした。また、OS が発行する I/O を監視、記録することで仮想化していないデバイスの状態を取得し、移動先のマシンで実際に I/O を発行することで、デバイス状態の復元した。

### 参考文献

1) Clark, Christopher., Fraser, Keir., Hand, Steven., Hansen, Jacob Gorm., Jul, Eric., Limpach, Christian., Pratt, Ian., Warfield, Andrew.: Live migration of virtual machines, *Proc. 2nd conference on Symposium on Networked Systems Design and Implementation(NSDI 2005)*, pp. 273–286 (2005)

2) Liu, Jiuxing., Huang, Wei., Abali, Bulent., Panda, Dhabaleswar K.: High performance VMM-bypass I/O in virtual machines, *Proc. annual conference on USENIX '06 Annual Technical Conference(ATEC 2006)*, pp.3–3 , (2006)

3) Huang, Wei., Liu, Jiuxing., Abali, Bulent., Panda, Dhabaleswar K.: A case for high performance computing with virtual machines, *Proc. 20th annual international conference on Supercomputing(ICS 2006)*, pp.125–134 , (2006).

4) Pan, Zhenhao., Dong, Yaozu., Chen, Yu., Zhang, Lei., Zhang, Zhijiao.: CompSC: live migration with pass-through devices, *SIGPLAN Not.*, Vol.47, pp. 109–120 (2012).

5) Kadav, Asim., Swift, Michael M.: Live migration of direct-access devices, *SIGOPS Oper. Syst. Rev.*, Vol.43, pp. 95–104 (2012).

6) Jeffrey, C., Mogul., Jayaram, Mudigonda., Jose, Renato, Santos., Yoshio, Turner.: The NIC Is the Hypervisor: Bare-Metal Guests in IaaS Clouds, (*HotOS 2013*), (2013)

7) Shinagawa, T., et al: BitVisor: A Thin Hypervisor for Enforcing I/O Device Security, *Proc. 2009 ACM SIGPLAN/SIGOPS international conference on Virtual execution environments (VEE 2009)*, pp. 121–130 (2009).