

平成 29 年度 卒業論文概要			
所 属	機械情報工学科	指導教員	光来 健一
学生番号	14237076	学生氏名	村岡 裕二
論文題目	仮想メモリを用いた VM マイグレーションの高速化		

1 はじめに

近年、クラウドサービスの一つとして、ユーザに仮想マシン (VM) を提供する IaaS 型クラウドが普及している。それに伴い、大容量のメモリを持つ VM が提供されるようになってきている。例えば、Amazon EC2 では 4TB のメモリを持つ VM が提供されており、ビッグデータの解析等に利用されている。VM はホストのメンテナンス等の際に別のホストにマイグレーションされるが、大容量メモリを持つ VM の移送先として十分な空きメモリを持つホストを確保しておくのはコスト面での負担が大きい。そこで、VM のメモリを複数の小さなホストに分割して転送する分割マイグレーション [1] が提案されている。この手法はマイグレーションを高速に行うことができるが、マイグレーション後に VM が他のホストにあるメモリを必要とした場合にはネットワーク転送が必要となりオーバーヘッドが大きい。一方、従来のマイグレーションにおいて、移送先ホストで仮想メモリを用いれば VM のメモリの一部は高速な SSD に格納されるため、マイグレーション後の性能低下を抑えることができる。しかし、従来の仮想メモリはマイグレーションとの相性が悪く、マイグレーション性能が低下するという問題があった。

本研究では、VM 専用の仮想メモリと分割マイグレーションの技術を組み合わせることにより高速な VM マイグレーションを実現するシステム VMemDirect を提案する。

2 大容量メモリを持つ VM のマイグレーション

IaaS 型クラウドでは大容量メモリを持つ VM が提供されるようになっており、ビッグデータの解析等に利用されている。VM が稼働しているホストをメンテナンスする際には、VM を停止させずに別のホストにマイグレーションすることで VM の実行を継続することができる。マイグレーションでは、移送先ホストにネットワーク経由で VM のメモリなどの状態を転送し、移送先ホストで VM を再開する。マイグレーション中に VM のメモリの内容が更新されると当該メモリのデータを移送先ホストに再送する。このように、マイグレーションを行うには、移送先ホストに VM のメモリよりも大きな空きメモリが必要となる。大容量メモリを持つ VM の場合、移送先として十分な空きメモリを持つホストを確保し続けることはコスト面で大きな負担となる。

そこで、VM のメモリを複数の小さなホストに分割して転送する分割マイグレーション [1] が提案されている。分割マイグレーションでは、VM 本体と可能な限りのメモリデータをメインホストに転送し、メインホストに入り切らないメモリデータはサブホストに転送する。マイグレーション後にサブホスト上に存在するメモリデータが必要になると、VM はサ

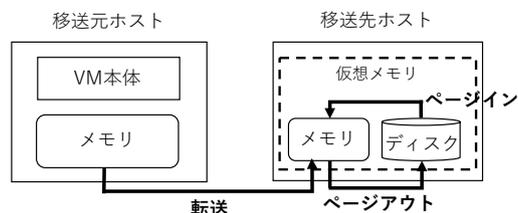


図1 仮想メモリを用いたマイグレーション

ブホストからメインホストにそのデータを転送 (ページイン) し、代わりに不要なメモリデータをサブホストに転送 (ページアウト) する。この処理はページングと呼ばれる。ページングの際にはネットワーク転送が行われるため、オーバーヘッドが大きくなり、性能を向上させるには高価な高速ネットワークが必要になる。また、ネットワークに障害が発生すると VM の実行を継続できなくなる。

一方、図 1 のように移送先ホストで仮想メモリを用いることにより、従来のマイグレーションを行うことも可能である。仮想メモリは VM のメモリの一部をディスクに格納する技術であり、VM がディスク上に存在するメモリデータを必要とした時には、当該データを物理メモリ上に転送してページインを行う。代わりに、物理メモリ上の不要なメモリデータをディスクに転送してページアウトを行う。近年の高速で安価な SSD を用いることにより、このページングのオーバーヘッドを削減することができる。しかし、従来の仮想メモリはマイグレーションとの相性が悪く、マイグレーションの性能が低下する。マイグレーション中に移送先ホストの物理メモリに空きがなくなるとページアウトが発生し、メモリの再送時にはページインが発生するためである。また、VM を実行する仮想化ソフトウェアのメモリもページアウトされる可能性が高く、マイグレーションの最終段階でページインが発生する原因となる。

3 VMemDirect

本研究では、VM 専用の仮想メモリと分割マイグレーションの技術を組み合わせることにより VM マイグレーションの高速化を実現するシステム VMemDirect を提案する。VMemDirect では移送先ホストの SSD 上に VM 専用のスワップファイルを作成し、仮想メモリとマイグレーションを密に連携させる。VM のメモリ以外はページングの対象としないため、仮想化ソフトウェアのメモリがページアウトされることによる性能低下を防ぐことができる。

3.1 直接メモリ転送

VMemDirect は VM のメモリを移送先ホストに転送する際に、従来の仮想メモリのページングを行うのではなく、物理メモリまたは SSD 上のスワップファイルのいずれかにメモリ

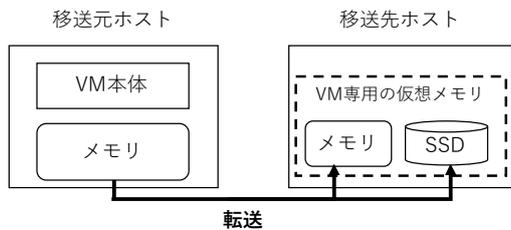


図2 VMemDirectによるマイグレーション

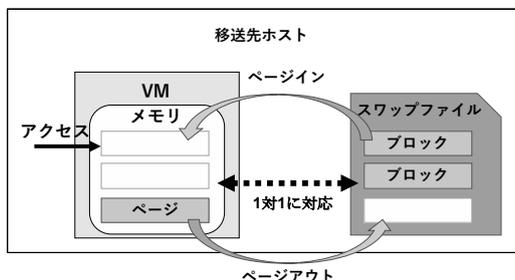


図3 VM専用の仮想メモリ

データを直接転送する。移送元ホストではマイグレーションを開始する際にメモリデータの格納先を決定し、メモリを再送する際も同じ格納先に転送する。メモリデータの格納先を移送先ホストに通知するために、転送するメモリデータの付加情報として送信する。これにより、マイグレーション中にページングが発生しないようにすることができる。物理メモリ上のデータはページアウトされることはなく、スワップファイルにあるメモリデータはページインされることはない。メモリの再送時には、スワップ領域のデータを直接上書きすることで更新する。

VMemDirectでは、移送元ホストにおけるVMのメモリアクセス履歴に基づいて、移送先ホストでのメモリデータの格納先を決定する。メモリアクセス履歴はCPUによって記録されるアクセス情報を利用して取得する。VMemDirectはメモリアクセス履歴を用いて、最近アクセスされたメモリ領域は今後もアクセスされると予測する。この予測に基づいて、アクセスされそうなメモリ領域は物理メモリへ、それ以外のメモリ領域はスワップファイルへ転送する。これにより、再送されるメモリデータの格納先が物理メモリである可能性が高くなり、再送時のオーバーヘッドを減らすことができる。

3.2 VM専用の仮想メモリ

VMemDirectでは仮想化ソフトウェアがVMごとにページングを行うため、VM単位でスワップファイルを作成する。スワップファイルはVMのメモリと同じサイズのスパースファイルと呼ばれる特殊なファイルとして作成し、図3のようにVMのメモリの各領域（ページ）とスワップファイルの各領域（ブロック）を1対1に対応づける。VMのメモリデータが物理メモリ上にない場合だけスワップファイルの対応するブロックにデータが格納され、それ以外のブロックは実データを持たない。

VMが物理メモリ上に存在しないメモリ領域にアクセスすると、仮想化ソフトウェアがそれを検出してページングを行う。まず、スワップファイルの対応するブロックを読み込み、VMの対応するメモリページにそのデータを書き込むことでページインを行う。同時に、スワップファイルの対応するブ

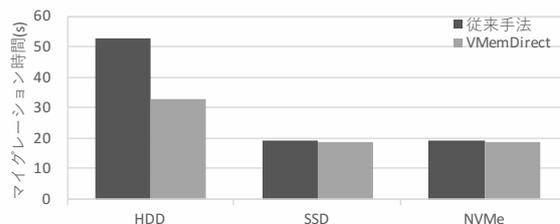


図4 マイグレーション時間

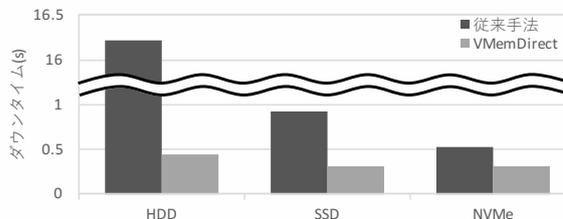


図5 ダウンタイム

ロックは空にする。次に、メモリアクセス履歴に基づいてアクセスされないことが予測されるメモリページを選ぶ。そのデータをスワップファイルに書き込み、VMのメモリページを削除することによりページアウトを行う。

4 実験

VMemDirectによるマイグレーションの性能を調べるために、マイグレーション時間とマイグレーション中にVMが一時停止する時間（ダウンタイム）を測定した。比較として、OSの仮想メモリを利用する従来手法を用いた。仮想メモリが用いるスワップ領域としてHDD、SSD、NVMeを使用した。実験には、Intel Xeon E3-1226 v3のCPU、8GBのメモリを搭載したマシンを2台用い、ギガビットイーサネットで接続した。VMには2GBのメモリを割り当て、移送先ホストの空きメモリを1GBにしてマイグレーションを行なった。

図4の実験結果より、HDDを用いた場合のマイグレーション時間は、VMemDirectにより37.5%短縮された。SSD、NVMeを用いた場合のマイグレーション時間は同程度となった。これはネットワーク速度がボトルネックとなったためと考えられる。ダウンタイムは図5のようになり、HDD、SSD、NVMeを用いた場合、従来手法と比較してそれぞれ97%、67%、42%短縮することができた。

5 まとめ

本研究では、VM専用の仮想メモリと分割マイグレーションの技術を組み合わせることによりVMマイグレーションの高速化を実現するシステムVMemDirectを提案した。VMemDirectは仮想メモリとマイグレーションを密に連携させることにより、マイグレーションの性能低下を抑える。今後の課題は、ページアウトの実装やメモリアクセス履歴の利用である。また、仮想メモリを用いているVMのマイグレーションにも対応する必要がある。

参考文献

- [1] M. Suetake, H. Kizu, and K. Kourai, Split Migration of Large Memory Virtual Machines, Proc. APSys 2016.