

複数の仮想マシン移送における移送コストの定量的調査

古藤 明音^{1,a)} 河野 健二^{1,b)}

概要：近年，クラウドコンピューティングにおいてオペレーティングシステム (OS) の仮想化技術が広く利用されている．特に OS の仮想化技術の一つである仮想マシン (VM) の Live migration を用いることで，クラウド環境における大規模なリソースを効率良く管理することができる．近年 Live Migration の研究は広く行われており，様々な移送手法が提案されているが，いずれの手法でも Live Migration の実行により VM のパフォーマンスが低下する問題がある．そこで，本研究ではクラウド環境を想定した移送環境において，live migration が引き起こすクラウドサービスの性能低下について定量的な調査をおこなった．調査の結果，live migration によるクラウドサービスの性能低下は移送する VM の種類や移送順序によって緩和することができることがわかった．特に，メモリ更新をあまりおこなわない VM を優先して移送すると移送時間を削減することができ，クラウドサービスにおける性能低下を緩和できることがわかった．また，メモリ更新を頻繁におこなう VM も移送する必要がある場合には更新頻度や更新範囲が小さい VM を優先して移送すると良いということがわかった．また，移送時間の大小が必ずしもアプリケーションの負荷を分散する速度と一致するとは限らないということがわかった．

キーワード：OS の仮想化技術，Live migration，サーバコンソリデーション

1. はじめに

近年，クラウドコンピューティングにおいて，オペレーティングシステム (OS) の仮想化技術が広く利用されている．OS の仮想化技術は，一台の物理サーバを複数の仮想的なサーバ (VM) として管理する機構を提供している．実際に Amazon EC2 [1] や Microsoft Azure [2] を始めとする主要なクラウドサービスも基盤技術に OS の仮想化技術を用いてサービスを用いている．OS の仮想化技術を用いたクラウド環境では，膨大な計算資源を VM に分割して提供することで，クラウド利用者の要求に応じた計算資源の提供を可能にしている．

OS の仮想化技術を基盤とするクラウド環境では，VM の live migration [3] を用いることでより柔軟な資源管理を実現することができる．Live migration とは，ネットワークのコネクションやアプリケーションの実行を維持したまま VM を移送する技術である．この技術を用いることで，クラウド環境では VM 上で稼働しているサービスを停止させることなく物理サーバのメンテナンスを行ったり，サーバ上で発生した突発的な負荷に追従した負荷分散をおこな

える．

しかしいずれの live migration 手法でも移送の実行により VM のパフォーマンスが低下する問題がある．これは live migration の実行に伴う資源消費により，移送を実行するホスト上で資源競合が発生するためである．live migration の実行には主に CPU 時間やネットワーク帯域を要する．そのため，例えばあるホストで増大した負荷を live migration によって緩和させようとしても，live migration の実行による資源消費が負荷を増大させてしまう．結果，移送する VM や (VMmg) や同ホスト上で稼働するその他の VM (VMco) では，負荷分散の実施によってパフォーマンスがさらに低下する．自身の研究 [4] より，live migration に伴う資源消費は VMmg や VMco のワークロードに応じて複雑に変化することがわかっている．特に，クラウドのようにワークロードの異なる様々な VM を稼働するホスト群を管理する環境では，live migration の稼働自体が複雑になる．例えば，複数の VM を異なるホストに移送することで負荷分散をおこなったり，特定のホストに複数の VM を同時に集約するなど，live migration が連続的または並列して実行することがある．こうした複数の VM 移送を伴う live migration は先に述べた live migration の性能変動を複雑化させ，予期せぬクラウドサービスのパフォーマンス低下を引き起こす原因となる．

¹ 慶應義塾大学
Keio University

a) koto@sslslab.ics.keio.ac.jp

b) kono@sslslab.ics.keio.ac.jp

そこで、本研究は複数の VM を移送する live migration がクラウドサービスに及ぼす影響について調査する。本研究により、live migration によるクラウドサービスの性能低下は移送する VM の種類や移送順序によって緩和することができることがわかった。特に、メモリ更新をあまりおこなわない VM を優先して移送すると移送時間を削減することができ、クラウドサービスにおける性能低下を緩和できることがわかった。また、メモリ更新を頻繁におこなう VM も移送する必要がある場合には更新頻度や更新範囲が小さい VM を優先して移送すると良いということがわかった。また、移送時間の大小が必ずしもアプリケーションの負荷を分散する速度と一致するとは限らないということがわかった。

本論文の構成を以下に示す。2 章では live migration が引き起こすクラウドサービスのパフォーマンス低下について代表的なシナリオを用いて紹介する。3.1 章では 2 章で紹介した代表的なシナリオについて、live migration が引き起こす性能低下を定量的に示す。4 章では関連研究について紹介する。そして 5 章において本研究のまとめを述べる。

2. Live migration によるクラウドサービスのパフォーマンス低下

Live migration は CPU 時間やネットワーク帯域を消費することから、VMmg や VMco のパフォーマンスを低下させることがある。本章では、自身の調査結果 [4] を踏まえ、live migration がクラウドサービスのパフォーマンスに及ぼす影響について説明する。

2.1 Live migration の性能変動

クラウド環境での live migration の実行によって引き起こされるサービスのパフォーマンス低下の大きさや傾向は、主に live migration により生じるダウンタイム、移送時間、資源消費によって変化することが知られている。ここでは、これら live migration の性能変動がクラウドの稼働においてどのような影響を持ち、どういった問題が発生するのかを述べる。

ダウンタイム: ダウンタイムは多くの live migration の研究において考慮される性能評価軸の一つである。Live migration では VM の稼働を維持しながら VMmg のメモリ情報を移送先に同期しているが、ある程度のメモリ情報が同期された後、VMmg の稼働場所を移送元ホストから移送先ホストへ切り替える必要がある。この切り替えの間、VMmg の稼働はサスペンドする必要があり、この時間をダウンタイムと定義している。クラウド環境では様々なサービスが絶えず稼働しており、なかにはミッションクリティカルサービスのような、稼働を停止できないサービスなどが稼働している可能性もある。こうした環境では、ダウン

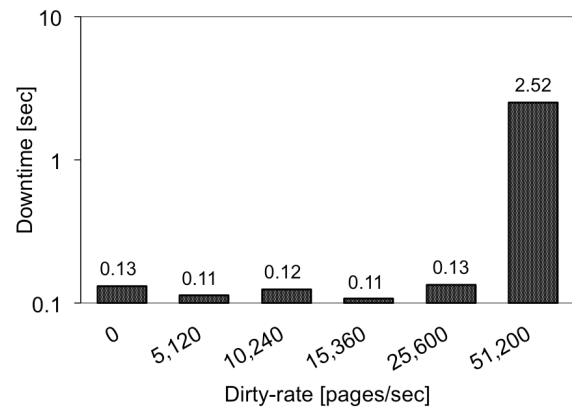


図 1 VMmg のメモリアクセスの違いによるダウンタイムの変化 (WSS: 64MB)

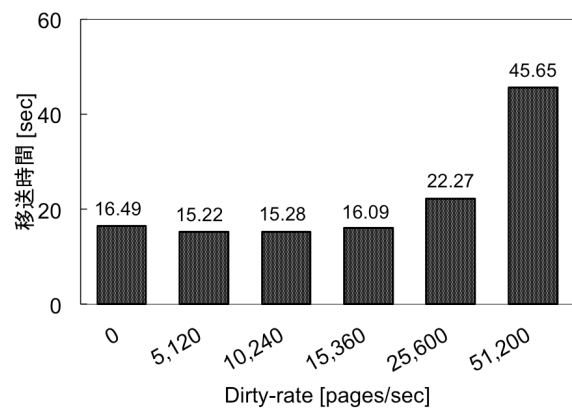


図 2 VMmg のメモリアクセスの違いによる移送時間の変化 (WSS: 64MB)

タイムの大きさがクラウドの品質に直接的に関わるため、できる限り緩和する必要がある。Live migration は VMmg が移送中に更新したメモリページの数 (Working set size: WSS) やメモリの更新速度 (Dirty-rate) によって VMmg の移送速度や移送にともなうメモリ転送の総量を調節している。そのため、クラウド環境では live migration は移送するクラウドサービスの稼働状況に応じて移送の性能が変動する。図 1 では VMmg が異なるワークロードを実行している生じたダウンタイムである。図より、live migration によって生じるダウンタイムは VMmg のワークロードによって最大で 2 秒以上異なっている。こうしたダウンタイムの変動はクラウドサービスの性能を大きく変動させる原因となる。

移送時間: ダウンタイムと同様に、移送時間も VMmg のワークロードによって変動することがわかっている (図 2)。クラウド環境では複数の物理サーバ群で構成される大規模な計算資源を管理しているため、クラウドの性能を向上す

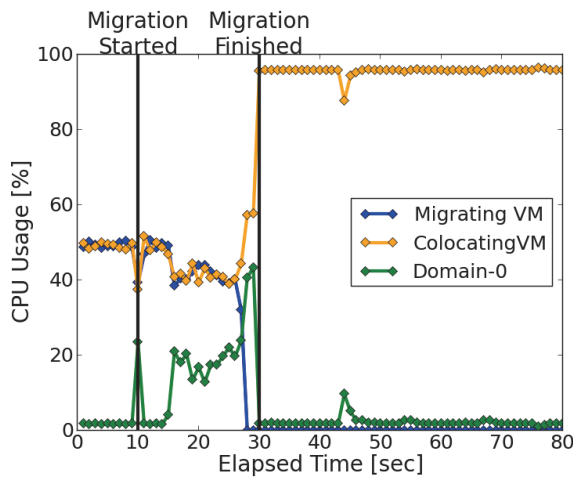


図 3 Live migration 実行による VMmg, VMco の CPU 使用量の変化 (WSS: 64MB, Dirty-rate: 15,360 pages/sec)

るためには大規模な計算資源を効率良く管理する必要がある。Live migration はこのような状況において、クラウドサービスが稼働する裏で計算資源の動的な管理を実現できる。例えば、クラウドサービスの要求する計算資源の量に応じて VM の配置を変更することで、クラウドで生じた負荷を動的に分散することができる。しかし、図 2 のように live migration にかかる時間が変動してしまうと、クラウドの負荷の変化に追従して負荷を分散できない可能性がある。

資源使用量: ダウンタイムや移送時間に加え、近年様々な研究で言及されているのが live migration の実行に伴う資源消費である。先でも述べたように、live migration は VMmg のメモリ情報を転送するために CPU 時間やネットワーク帯域を消費する。これにより、特に負荷分散を実行する際に計算資源が不足している物理サーバが live migration を実行する場合、live migration によって負荷がより深刻になる可能性がある。図 3 はある物理マシンにおいて稼働していた 2 台の VM (migrating VM: VMmg, colocating VM: VMco) において、CPU 時間の資源要求が増大し資源競合が発生したため live migration によって負荷分散を実施しようとした際の VMmg, VMco, 移送プロセス (Domain-0) の CPU 使用量の変化である。図より、VMmg と VMco に割り当てている CPU 時間は live migration 開始 (10 秒後) から終了 (30 秒後) までの間、移送開始前に比べ低下していることがわかる。これは、live migration を実行している domain-0 が VMmg を転送するために 20-40% の CPU 時間を消費したためである。実際に、VMmg と VMco は移送中に 10% 以上の CPU 時間が低下しており、負荷分散のために live migration を実行した結果、負荷が深刻化した。このような負荷の深刻化は移送時間やダウンタイムの変動によりさらに複雑な形で発生する。ここでは、クラウドサービスにおいて予期せぬパ

フォーマンス低下が発生する原因となる。

2.2 クラウド環境での live migration の利用例

2.1 で述べたように live migration によってクラウドサービスのパフォーマンスは複雑に変化するが、クラウドの稼働状況は様々であり稼働時間の経過と共に変化していくことから、包括的に調査することは難しい。そこで本研究では、クラウド環境において想定される live migration の利用例として負荷分散とマシンメンテナンスにおける VM の移送を想定した際に発生するサービスのパフォーマンス低下について調査する。本章では、それぞれのシナリオについて想定される移送環境や移送方法について紹介する。

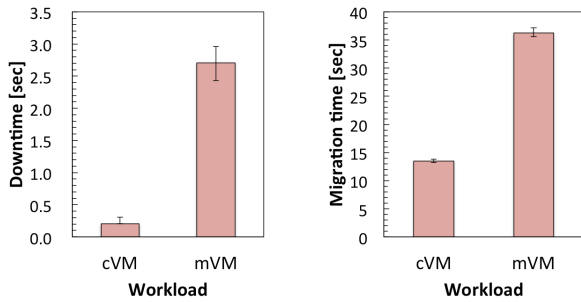
シナリオ 1: 負荷分散 Live migration により負荷分散を実行する場合、移送元ホストでは資源競合が発生している状態である。2.1 で述べたように、live migration の稼働は VMmg の稼働によって変化する。そのため、どの VM を移送するかで live migration によって得られる負荷分散の効果は変化する可能性がある。特に、live migration に要する移送時間は VMmg に、負荷の増大の仕方は VMmg を含むホスト全体の稼働によって変動する可能性がある。同シナリオでは、急激な負荷に追従した負荷分散が実行できるような、高速な live migration が必要とされる。また、それに加え、クラウドサービスの最悪性能を保証できるよう、負荷の小さい live migration が必要とされる。

シナリオ 2: マシンメンテナンス 負荷分散の例とは異なり、マシンメンテナンスを目的とした live migration では、メンテナンスを実施する当該物理マシン上で稼働する全ての VM を移送する必要がある。この時、VM の移送順序により、移送元・移送先ホストの稼働状況の変化は異なる。そのため、移送順序を適切に選択しなければ移送元、移送先のホストで予期せぬサービスの性能低下が発生する可能性がある。例えば、一部の VM で発生するダウンタイムや移送時間が長くなり、一部のサービスでのみパフォーマンスが著しく低下する可能性がある。また、特に移送時間が長くなってしまった場合、live migration による資源消費が発生する時間も長くなることから、VMco でもパフォーマンスが低下する可能性がある。そのため、同シナリオではダウンタイム、移送時間、資源使用がより緩和できる移送順序を選択することが重要である。

3. 調査

3.1 実験環境

調査には 2 台の移送用ホストマシンと 1 台の NFS を用いた。VM は計 4 台用意し、全ての VM を移送元ホストから移送先ホストへ live migration する。3 台の物理マシンにはいずれも Xeon 2.67GHz X5650 (6 コア), 32GB のメモリを搭載する Dell PowerEdge T410 を用いた。これらの物理マシンは Cisco Catalyst 3750G ネットワークス



(a) ダウンタイム (b) 移送時間

図 4 2 台の VM によって資源競合を発生しているホストマシンにおいて、どちらか一方の VM を live migration した場合に得られた移送の性能結果

イチとギガビットイーサネットを用いて接続している。VM はいずれも 512 MB のメモリと VCPU 1 つを搭載している。全ての物理マシン, VM 上では Linux 3.19.3 カーネルを稼働させた。仮想化環境には Xen-4.4.1 を用い, live migration は同仮想化環境に実装されている pre-copy live migration を用いた。

VM で稼働させるワークロードは CPU-intensive, Memory I/O-intensive の二種類を用意した。二つのワークロードはいずれも自作の C プログラムによるものである。CPU-intensive では, カウンターをインクリメントする無限ループを実行するプログラムであり, 実行中は単位時間あたりのカウント数をスループットとして出力する。Memory I/O-intensive では, 460 MB のメモリを malloc() 関数で確保し, そのメモリを 1 ページずつ更新し続けるプログラムである。このプログラムは単位時間あたりに更新したメモリページ数をスループットとして出力する。

3.2 実験結果

3.2.1 負荷分散におけるサービスの性能低下

この実験では, 2 台の VM を用い, それぞれ CPU-intensive, Memory I/O-intensive を稼働させた。同実験では物理マシンの CPU コア数を 1 つに設定し, CPU-intensive VM (cVM), Memory I/O-intensive VM (mVM) Domain-0 がコアを共有するように割り当てた。両 VM はいずれもそれぞれの 1 台で稼働している状態では 100% の CPU 時間を使用している。この時, 2.2 で述べた負荷分散のための移送を模し, 2 台の VM のうちどちらかを移送することで移送時間, ダウンタイム, 資源使用がどのように異なるのか定量的に調査した。以降の実験はいずれも 5 回ずつ実施している。

図 4 は実験により得られた移送時間とダウンタイムである。この図より, 1 台の VM を移送することで負荷分散を実施する場合には cVM を移送した方がダウンタイム, 移送時間共に緩和することができ, サービスの性能低下を緩和しながら負荷分散を実施できることがわかる。これは, 2.1 で述べた VMmg のワークロードによる live migration

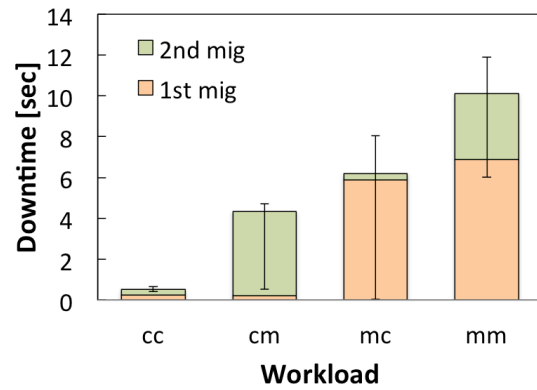


図 5 資源競合を引き起こしている 4 台の VM のうち, 2 台を live migration した場合のダウンタイム

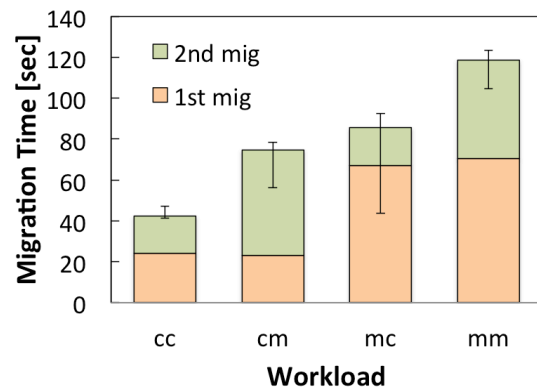


図 6 資源競合を引き起こしている 4 台の VM のうち, 2 台を live migration した場合の移送時間

の性能変動によるものである。特に, cVM と mVM はいずれも 100% の CPU 時間を使用しているが, cVM ではメモリの更新を行わないのに対し, mVM では VM に割り当てているほとんどのメモリを高速に更新し続けている。そのため, mVM を移送する際には, メモリ情報の転送中に大量のメモリページが更新されてしまい, 移送ホスト間でメモリ情報を同期するために大量のメモリ転送が発生する。これより, 負荷分散のために 1 台の VM を移送する場合にはメモリ更新をあまりおこなわないサービスを優先して移送するべきであると言える。

次に, 負荷分散によって 2 台以上の VM を移送する場合の移送の性能変動について調査した。この実験では cVM と mVM を 2 台ずつ, 計 4 台用い, 全ての VM を移送元ホストで稼働させた。この時, 4 台のうちいずれか 2 台の VM を移送した。同実験で得られたダウンタイム, 移送時間の結果を図 5, 図 6 にそれぞれ示す。図 6 より, 2 台の VM を移送することで負荷分散が終了するまでにかかる時間は cVM 2 台を移送した時 (cc) で 42.37 秒, mVM 2 台を移送した時 (mm) で 118.81 秒となり, 最大で 76.44 秒異なることがわかった。また, cVM, mVM を一台ずつ移送した場合には, cVM, mVM の順で移送した場合 (cm) と, mVM, cVM の順で移送した場合とで 11.68 秒, 移送

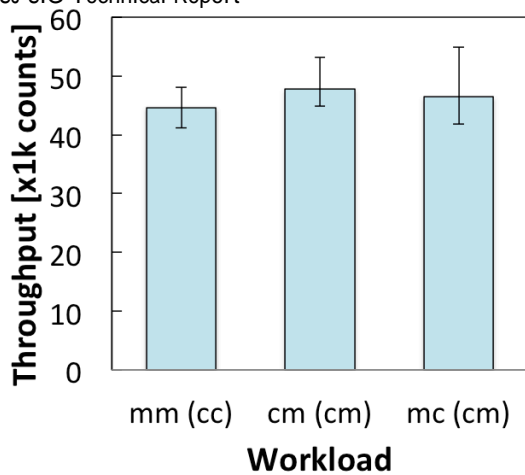


図 7 資源競争を引き起こしている 4 台の VM のうち、2 台を live migration した場合の cVM 1 台あたりの平均スループット

時間の合計が異なることがわかった。これは、同じワークロードを稼働させている VM でもホスト全体の資源使用量が異なる場合には移送の性能が変化するためである。特に、cm の順で移送した場合には mVM の移送が実行される間、移送元ホストでは 3 台の VM によって資源競争が発生している。これに対し、mc の順で移送した場合、mVM の移送が実行されている間に移送元ホストで発生する資源競争は 4 台の VM によって引き起こされている。そのため、mc の順で移送を行った場合には、の移送時に live migration が移送に十分な計算資源を確保できず移送が長期化してしまった。cm の順で移送した場合にはの移送時に移送の長期化が発生するが、図 4 からわかるように、cVM と mVM とでは 1 台の移送に要する時間が大きくことなることから、mc での mVM の移送に比べ、大きな移送の長期化が発生しなかった。

2 台の VM の移送により発生するダウンタイムの合計も、移送する VM の種類や順序によって異なることがわかった。図 5 より、cc, mm で移送を実行した場合、2 台の VMmg に発生したダウンタイムの合計は 9.56 秒異なった。また、cm, mc とで移送順序を変えた場合でも 2.24 秒の差が生じた。この時、mVM において発生したダウンタイムは cm, mc それぞれで 5.89, 4.13 秒となり 1 秒以上の差があることがわかった。こうした移送時間、ダウンタイムの差はクラウドサービスに発生する性能低下の大きさや規模を変動させる原因となる。今回の実験より、ダウンタイム、移送時間から引き起こされるサービスの性能低下を緩和するためには cVM を優先して移送した方が良いということがわかった。

次に、実際に 2 台の VM を移送することで移送元ホストに残された残りの 2 VM では、負荷分散によりスループットがどのように上昇したのかを比較する。図 7, 図 8 に、live migration を実行する 30 秒前から 100 秒間の間に移送されなかった VM で得られたスループットの総計を示す。それぞれの図において x 軸は [移送した VM](移送元に残っている VM) を表している (例: cm(cm) は 2 台の

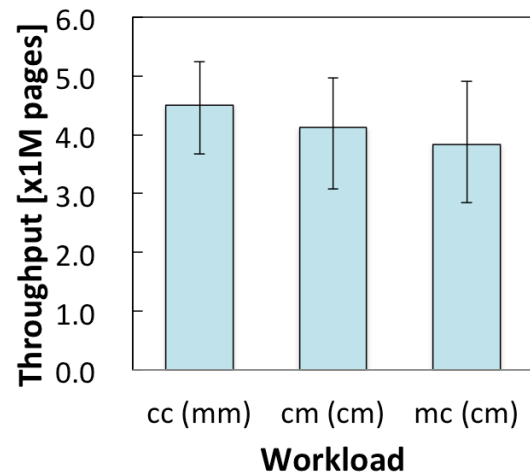


図 8 資源競争を引き起こしている 4 台の VM のうち、2 台を live migration した場合の mVM 1 台あたりの平均スループット

VMmg を cm の順で移送し、移送元ホストには cVM と mVM が 1 台ずつ残っている)。二つの図より、移送時間やダウンタイムは移送する VM の種類や移送順序によって変化することがわかったが、スループットの変動はワークロードの性質によって異なることがわかった。特に図 2 より、2 台の VM の移送に要する時間は cVM を優先した移送ほど短くなっていったが、cVM のスループットは移送した VM の種類や順序の影響をあまり受けていないことがわかる (図 7)。実際に、最もスループットが低かったのは 2 台の mVM を移送した時であったが、これは最もスループットが高かった cm(cm) 時と比べ 6.59% 低かった程度であった。一方、mVM のスループットは移送する VM の種類や順序によって大きく変化していることがわかる (図 8)。実際に、最もスループットが高かった cc(mm) では mc(cm) 時と比べ 14.74% 高いスループットを得た。これより、移送時間による負荷分散の速度は必ずしもクラウドサービスにかかる負荷を分散する速度に比例するとは限らないことがわかった。

4. 関連研究

Live migration 中のネットワークトラフィックを削減する手法が広く研究されている [5], [6], [7], [8], [9]。MECOM [5] は移送 VM のメモリページの特徴を考慮したメモリ圧縮機能を用いて転送するメモリの総量を削減した。Svard らは移送 VM のメモリページのうち、頻繁にアクセスされるメモリページを一つのまとまりとして扱い、それらを排他的論理和によって圧縮することでメモリの総量を削減した [6]。CR/RT-Motion [7] は移送 VM の動作を追跡したログのみを移送先へ転送し、移送先ではそのログを用いて VM を再現した。これらの研究は移送プロセスのリソース消費の削減に着目した研究であり、VMmg や VMco のパフォーマンス変化や移送中の性能変化を考慮していない。

Live migration の実行をモデル化することでクラウドサービスのパフォーマンス低下を緩和する研究がされている。Breitgand ら [10] は live migration の実行に必要な

ネットワーク帯域を移送のステップごとに決定することで移送コストをモデル化した。Lim ら [11] は live migration の動作を移送元と移送先の移送プロセスの組として扱い、移送コストをモデル化した。これらの研究は代表的なワークロードを実行する際には効果的なモデルであるが、クラウド環境で発生する突発的な負荷増加などの予測が難しいワークロードについては考慮していない。

Hu ら [12] は live migration のパフォーマンスをいくつかの軸を用いて評価することで live migration の稼働を動的に評価するテスト用フレームワークを実装した。この研究では VMco のパフォーマンスへの影響については考慮していない。本研究では VMmg と VMco の両方について定量的な調査をおこなっている。

5. まとめ

VM の live migration はクラウド環境におけるコンピュータリソースを柔軟に管理するのに有用な技術である。live migration は広く研究されており様々な移送手法が存在するが、いずれの手法においても live migration の実行により移送 VM や移送ホスト上で稼働するその他の VM のスループットが低下する問題が起こる。こうした live migration によるクラウドサービスの性能低下の大きさや種類は VM のワークロードや物理マシンの計算資源の使用状況によって変化する。そこで、本研究ではクラウド環境において live migration が実行される移送環境を想定し、複数の VM を移送した際に発生するクラウドサービスの性能低下について定量的に調査した。調査の結果、移送によるクラウドサービスの性能低下を緩和するにはメモリよりも CPU 時間を消費している VM を優先して移送した方が良いということがわかった。特に、複数の VM を移送する環境において、メモリを消費する VM も移送しなければならない場合には、先に CPU 時間を消費する VM を移送した方がクラウドサービスに及ぼす影響は小さくなるということがわかった。また、移送時間の大小が必ずしもアプリケーションの負荷を分散する速度と一致するとは限らないということがわかった。

参考文献

- [1] : Amazon Elastic Compute Cloud (Amazon EC2).
- [2] : Microsoft Azure.
- [3] Clark, C., Fraser, K., Hand, S., Hansen, J. G., Jul, E., Limpach, C., Pratt, I. and Warfield, A.: Live Migration of Virtual Machines, *Proceedings of the 2nd Conference on Symposium on Networked Systems Design and Implementation (NSDI '05)*, pp. 273–286 (2005).
- [4] Koto, A., Kono, K. and Yamada, H.: A Guideline for Selecting Live Migration Policies and Implementations in Clouds, *Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on*, pp. 226–233 (online), DOI: 10.1109/CloudCom.2014.36 (2014).

- [5] Jin, H., Deng, L., Wu, S., Shi, X. and Pan, X.: Live Virtual Machine Migration with Adaptive Memory Compression, *Proc. of 2009 IEEE Int'l Conf. on Cluster Computing*, pp. 1–10 (2009).
- [6] Svård, P., Hudzia, B., Tordsson, J. and Elmroth, E.: Evaluation of Delta Compression Techniques for Efficient Live Migration of Large Virtual Machines, *Proceedings of the 7th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE '11)*, pp. 111–120 (2011).
- [7] Liu, H., Jin, H., Liao, X., Hu, L. and Yu, C.: Live Migration of Virtual Machine Based on Full System Trace and Replay, *Proceedings of the 18th ACM International Symposium on High Performance Distributed Computing (HPDC '09)*, pp. 101–110 (2009).
- [8] Jo, C., Gustafsson, E., Son, J. and Egger, B.: Efficient Live Migration of Virtual Machines Using Shared Storage, *Proceedings of the 9th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments*, pp. 41–50 (online), available from (<http://doi.acm.org/10.1145/2451512.2451524>) (2013).
- [9] Chiang, J.-H., Li, H.-L. and Chiueh, T.-c.: *Introspection-based Memory De-duplication and Migration*, Vol. 48, No. 7, pp. 51–62 (2013).
- [10] Breitgand, D., Kutiel, G. and Raz, D.: Cost-Aware Live Migration of Services in the Cloud, *Proc. of Workshop on Hot Topics in Management of Internet, Cloud, and Enterprise Networks and Services* (2011).
- [11] Lim, S.-H., Huh, J.-S., Kim, Y. and Das, C. R.: Migration, Assignment, and Scheduling of Jobs in Virtualized Environment, *Proceedings of the 3rd USENIX Conference on Hot Topics in Cloud Computing, HotCloud'11*, pp. 2–2 (2011).
- [12] Hu, W., Hicks, A., Zhang, L., Dow, E. M., Soni, V., Jiang, H., Bull, R. and Matthews, J. N.: A Quantitative Study of Virtual Machine Live Migration, *Proceedings of the 2013 ACM Cloud and Autonomic Computing Conference (CAC '13)*, pp. 11:1–11:10 (2013).