# 移動通信機能NEMO BSを用いた ゲスト計算機のライブマイグレーション

**効率のよい仮想環境を構築するには、仮想計算機の流動性を確保し柔軟に管理する仕組みが必要です。
ここでは、IPv6ルータに移動通信機能を追加するNEMO BS技術を利用した、
仮想計算機の移動方法を示し、その実験結果を考察します。
NEMO BSを使うことで、ネットワークセグメントを越えた仮想計算機の移動が可能になります。**

## 3.1 背景

計算機の処理能力、ネットワークやデータストア等の技術進歩により、個々の計算機の能力は飛躍的に進歩しています。一方、計算機単体の速度向上は今後鈍化するという見方もあり、近年は複数の計算機をひとつの計算機資源として取り扱うクラウド技術が注目されています[*1][*2]。また、ひとつの計算機資源を仮想的に分割し、複数の異なる計算機として利用する仮想計算機技術も長年研究されています[*3]。一見異なる方向を目指すように思えるこれらの技術ですが、互いを補うことでより効率よく計算機資源を活用できるものになると考えられています。たとえば、Amazonが提供するEC2[*4]は、複数の計算機資源を組み合わせてひとつのサービスにするクラウド環境を提供していますが、ひとつひとつの計算機資源には仮想計算機技術が用いられています。1台の物理的な計算機という大きな単位での資源管理をやめ、小さく切り出した仮想計算機をひとつの単位として利用することで、効率がよく柔軟な計算機資源の利用を実現し、結果的に仮想計算機に分割したことによる負荷を上回る利便性を提供しています。

このような仮想環境が発展していくためには、仮想計算機を柔軟に管理する仕組みが重要です。クラウド環境からの要求に従って、必要な場所に必要な量の仮想計算機を配置できることが、全体の性能向上や資源の効率よい利用に貢献するからです。ここでは、仮想計算機の再配置の仕組みに注目し、稼働中の仮想計算機を異なるセグメント(オフリンクセグメント)に移動させる手法を提案します。

## 3.2 ライブマイグレーションの課題

現在、多くの仮想計算機技術が実用レベルで提供されています。その中には仮想計算機を、その親となっている計算機(ホスト計算機)から別のホスト計算機に移動させる機能を提供しているものもあります。VMwareのVMotionやXen[*5]のLive Migrationがこの代表例です。以後、ここではホスト計算機から切り出された仮想計算機を「ゲスト計算機」、ゲスト計算機をホスト計算機間で移動する機能を「ライブマイグレーション」と呼びます。ライブマイグレーション機能を利用すると、稼働中のゲスト計算機をほとんど停止することなく別の機材に移動できます。ただし、現在提供されている機能には、移動元のホスト計算機と移動先のホスト計算機が同一のセグメントに属していなければならないという制限があります。

この制限は、ゲスト計算機に提供されているネットワーク構成方式によるものです。ホスト計算機とゲスト計算機の関係は対等ではありません。通常、ホスト計算機がすべての資源を管轄し、ゲスト計算機にその一部を配分します。ゲスト計算機をネットワークに接続す

＊1　Aaron Weiss. Computing in the clouds. *netWorker*, Vol. 11, No. 4, pp. 16-25, December 2007.
＊2　Brian Hayes. Cloud computing. *Communications of the ACM*, Vol. 51, No. 7, pp. 9-11, July 2008
＊3　Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, et al. Xen and the art of virtualization. In *SOSP '03: Proceedings of the nineteenth ACM symposium on Operating systems principles*, pp. 164-177. ACM, 2003
＊4　Amazon. Amazon Elastic Compute Cloud (Amazon EC2), October 2009. http://aws.amazon.com/ec2/
＊5　Citrix, October 2009. http://www.xen.org/

る場合、図-1に示すようなネットワーク構成が用いられます。図-1の構成(a)では、ゲスト計算機はホスト計算機と同じネットワークに、ホスト計算機が提供する仮想スイッチを経由して接続されます。構成(b)でも、ホスト計算機とゲスト計算機の間に仮想スイッチが設けられます。構成(a)と異なり、この場合、ホスト計算機はゲスト計算機の上流ルータとしても機能します。図-1からも明らかなように、ゲスト計算機のネットワーク構成はホスト計算機に大きく依存します。ライブマイグレーションの実行時には、ゲスト計算機自体の動作環境は変更しません。結果的に、構成(a)でしかライブマイグレーションは実現できません。構成(b)では、仮想スイッチに割り当てられるアドレスがホスト計算機によって異なります。これでは、ゲスト計算機を移動した後、ネットワーク環境を適切に変更しない限り通信を継続できません。また、構成(a)であっても、移動元と移動先のホスト計算機が異なるセグメントに接続している場合、同様の問題が発生します。

ライブマイグレーションによって、仮想計算機が1つのホスト計算機に集中することは回避できますが、その機能は同一セグメント内での移動に限られます。他のセグメントに資源に余裕のあるホスト計算機が配置されていても、それを活用することはできません。また、ゲスト計算機を利用者により近いホスト計算機に移動させるような、性能向上のための資源再配置もできません。

## 3.3 NEMO BSの概要

NEMO BS (Network Mobility Basic Support)*6は、IPv6ルータに移動通信機能を追加する仕様です。NEMO BS環境では、NEMO BSに対応したMR (モバイルルータ)が固定のネットワークプレフィックスであるMNP (モバイルネットワークプレフィックス)を管理します。MRが提供するネットワークに接続したIPv6ホストは、MNPの範囲にある固定アドレスを利用します。MRはインターネット上のさまざまなセグメントに接続し、接続先のネットワーク環境に応じてインターネットへの接続性を確保します。このとき、MRが管理するMNPは変化しません。MR配下のホストは、MRの位置にかかわらず常に同じネットワーク環境を維持できます。

この機能は、MRと対をなして動作するHA(ホームエージェント)によって実現されています(図-2)。MRは移動先のネットワークで、ネットワーク環境に応じたアドレス(気付アドレス)を取得し、HAとの間に双方向IPv6 over IPv6トンネルを確立します。MNP内のノードで発生したトラフィックは、このトンネルを使っていったんHAに送られ、HAから通信相手に転送されます。一方、MNP内部ノード宛のトラフィックは、HAでいったん受け取られ、トンネルを介してMRに配送された後、最終宛先のホストに転送されます。

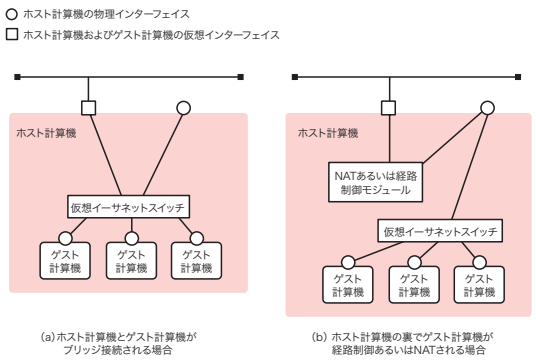


図-1　ゲスト計算機のネットワーク構成

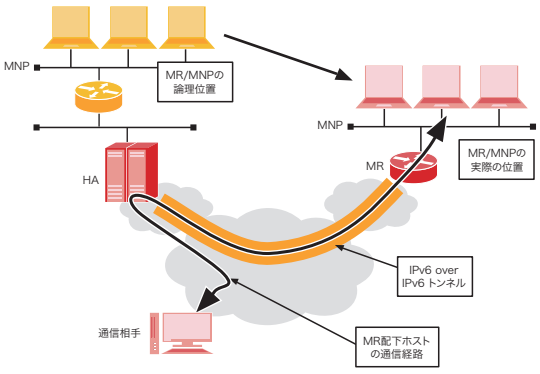


図-2　NEMO BSの動作概要

*6　Vijay Devarapalli, Ryuji Wakikawa, Alexandru Petrescu, and Pascal Thubert. *Network Mobility (NEMO) Basic Support Protocol*. IETF, January 2005. RFC3963

## 3.4 設計

ライブマイグレーションが同一セグメント内での移動に制限される理由は、ゲスト計算機のネットワーク環境がホスト計算機に依存しているためです。つまり、ゲスト計算機のネットワーク環境がホスト計算機によらず一定になるように設計すれば、異なるセグメントに接続しているホスト計算機にもゲスト計算機を移動できるようになります。

ここでは、ゲスト計算機のネットワーク環境を一定に保つ方法として、IPモビリティ技術を用いる方法を提案します。この方法には、ゲスト計算機自体がMobile IPなど[*7][*8]のホスト移動通信機能を備える方法と、ホスト計算機がNEMO BSなどを活用して固定ネットワークをゲスト計算機に提供する方法が考えられます。前者では、ゲスト計算機の改変(IPモビリティ機能の導入)が必要ですが、ゲスト計算機単体での移動が可能になり、計算機資源の細かな制御が期待できます。後者の方法では、これまで利用してきたゲスト計算機を無変更で継続利用できる代わりに、ゲスト計算機の移動とNEMO BSのMRとしてホスト計算機の移動が同期しなければならないという制限が生じます。今回は、既存のシステムで用いられているゲスト計算機をそのまま利用し続けるというシナリオを前提として、後者の方法に注目します。

システムの設計は、ホスト計算機の資源管理の方法によって異なる可能性があります。ここでは、Xenが提供する仮想計算機環境を用いた設計を提示していますが、他のシステムでも大きな変更は必要ないと思います。図-3がその概要です。構成は、図-1の構成(b)を拡張したものになります。ホスト計算機がMRとしての機能を持ち、インターネット接続用と、MNP提供用の2つのインターフェースを提供します。MNP接続用インターフェースは、物理的なインターフェースではなく仮想インターフェースです。仮想インターフェースは、ホスト計算機が提供するゲスト計算機用の仮想スイッチに接続されています。ホスト計算機の仮想インターフェース、および仮想スイッチに接続されたゲスト計算機のインターフェースには、MRが管理する固定アドレスが割り当てられます。NEMO BSの機能により、ホスト計算機が物理的にどこに接続されていても、MNP内のアドレスが変化することはありません。

ゲスト計算機をライブマイグレーションする環境では、複数のホスト計算機がネットワーク上に配置されます。これらのホスト計算機には、MRとして同じMNPが設定されますが、稼働中のゲスト計算機を有していないホスト計算機はMRとして動作しません。ゲスト計算機を移動する場合、まず通常のライブマイグレーションの手順を用いて、移動先のホスト計算機にゲスト計算機を移動します。この時点では、まだゲスト計算機はネットワークから切り離された状態です。その後、移動元のホスト計算機が移動先のホスト計算機に移動完了通知を送り、自分自身のNEMO BS機能を停止します。移動完了通知を受けたホスト計算機は、HAに対して現在位置を登録し、NEMO BSのMRとしての動作を開始します。HAへの登録が完了した段階で、ゲスト計算機が接続している仮想スイッチのMNPが有効となり移動が完了します。

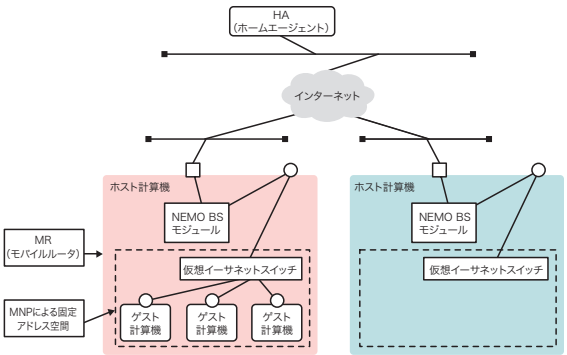


**図-3　NEMO BSを用いたゲスト計算機の移動**

＊7　Basavaraj Patil, Phil Roberts, and Charles E. Perkins. *IP Mobility Support for IPv4*. IETF, August 2002. RFC3344
＊8　David B. Johnson, Charles E. Perkins, and Jari Arkko. *Mobility Support in IPv6*. IETF, June 2004. RFC3775.

## 3.5 実験による検証

提案した設計が実現可能であることを確認するため、プロトタイプを実装して稼働実験を行いました。実験環境は、HA、NEMO BS機能を備えMRとして動作する2台のXenホスト計算機、テスト用のストリーミングデータを受信する計算機と、移動の指示を出す制御用計算機の合計5台で構成しました。Xenホスト計算機には1台のゲスト計算機を構築し、ストリーミングサーバとして動作させました。

実情に近い環境で実験するため、これらの機材は実際のインターネット環境上に配置しました。HA、およびMRが利用する固定ネットワークはInterop Tokyo 2009[*9]のために構築された運用ネットワークの一部に設置し、Xenホスト計算機とストリームデータ受信ノードはIIJのネットワーク上に配置しました。それぞれの機材はインターネットを介して相互接続しています。図-4に実験ネットワークの概要を示します。

ゲスト計算機(ストリーミングサーバ)からは、15Mバイト、520Kビット/秒のMPEG4ストリームデータをUDPを使って継続的に送信しました。また、今回、ゲスト計算機のライブマイグレーション開始指示、およびホスト計算機のNEMO BS機能の停止作業と開始作業は、別途用意した制御用計算機からsshを用いて行いました。制御用計算機から、ゲスト計算機のライブマイグレーションコマンドを遠隔実行し、ライブマイグレーションが完了した直後にホスト計算機のNEMO BS移動処理を実行しています。この操作を5分間隔で繰り返し、2台のホスト計算機間でのゲスト計算機の移動を繰り返しました。

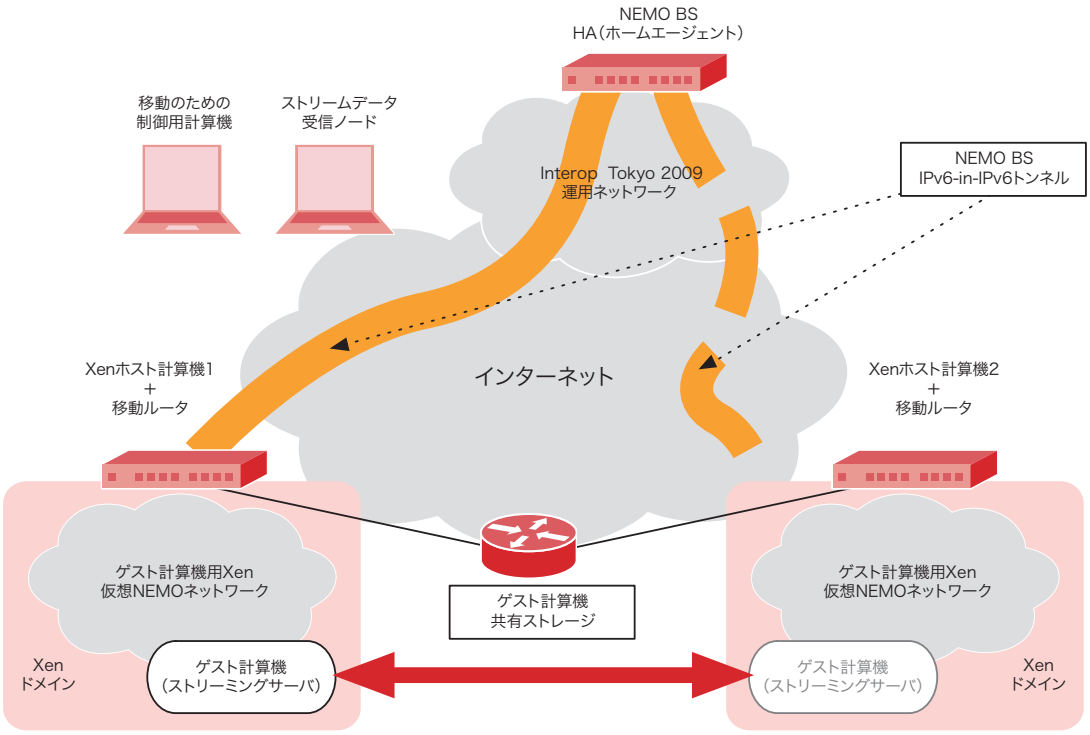


**図-4　実験ネットワークの概要**

*9　Interop Tokyo 2009, June 2009. http://www.interop.jp/

## 3.6 評価と課題

ホスト計算機のインターネット側インターフェースでストリームトラフィックをモニタした結果を図-5に示します。5分ごとにゲスト計算機がライブマイグレーションされ、トラフィックグラフが2台のホスト計算機の間を移動していることが確認できます。

ライブマイグレーションでは、仮想計算機のデータを移動先のホスト計算機に複製している間も、移動元の仮想計算機が動作し続けます。切り替え時の停止時間はわずか(数百ミリ秒)です。ただし、今回の構成ではネットワーク層の技術を用いてホスト計算機を移動させたため、ゲスト計算機の移動後にホスト計算機の移動処理も必要でした。トラフィックデータを詳細に調べたところ、一方のホスト計算機がストリーミングデータの配信を停止した後、移動先のホスト計算機がストリーミングデータを送信し始めるまでに6秒程度の時間がかかっていることが判明しました。

ホスト計算機が接続するネットワークでのルータ広告の間隔は3 ～ 4秒でしたので、ホスト計算機はおよそ2秒間でルータ広告を受信できる環境にありました。NEMO BSの処理の一部である、気付アドレスの重複確認に1秒間かかることを含めても、この環境では3秒程度で移動処理が完了するはずです。今回、それ以上に時間がかかっている理由として次の2点が考えられます。

1. 制御計算機からの遠隔操作スクリプト実行によるオーバヘッド
2. 通常の移動とは異なる手順

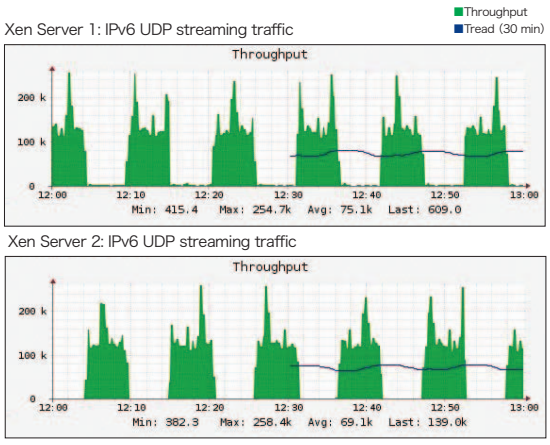


**図-5　ストリームトラフィックの推移**

NEMO BS機能の停止と開始をsshによる遠隔スクリプトで実行したため、TCP接続のための時間、ホスト計算機でのプロセス生成時間、遠隔コマンドによるNEMO BSデーモンプログラムの制御などの時間が、通常の運用と比較して余計に必要となっています。また、本来ならば単一の移動ノードで実行される移動処理が、今回は2つの異なる移動ノードで協調実行されています。つまり、一方のMRがHAへの登録解除処理を完了した後に、もう一方のMRが登録処理を実行しています。同一MRで移動処理を実行する場合、新しい情報の登録時に古い情報の登録を解除する必要はなく、連続した更新になります。今回は、解除のために余計な時間がかかっています。

## 3.7 考察

今回の実験は、2つの技術(NEMO BSとXen仮想化)の組み合わせ検証となっています。次に、検証の過程で得た知見を述べます。

### 3.7.1 ネットワークストレージの問題

仮想計算機環境を提供する仕組みでは、ゲスト計算機のストレージ提供方法として、ホスト計算機のストレージの一部を提供する方法と、ネットワークストレージを提供する方法の2種類が考えられます。ライブマイグレーションを利用する場合、ゲスト計算機を収容するホスト計算機が変わるため、ストレージはネットワーク上に配置しなければなりません。現在想定されている利用方法では、ゲスト計算機が同一セグメントを越えて移動することがないため、ネットワークストレージを利用しても移動前後に大きな環境の変化は生じません。しかし、今回の提案のようにセグメントを越えてゲスト計算機を移動する場合、ネットワークストレージへの到達性、応答性などが問題になる可能性があります。

1つの解決策として、外部ストレージを用いずに、すべてをメモリ上で処理するゲスト計算機環境を構築する方法が考えられます。しかし、この方法ではメモリ上に保持するデータ量が増えるに従って、移動完了までの時間が増加します。また、その構造上、大量のデータを扱うことが困難になります。

別の方法として、インターネット上の複数地点から効率よく透過的に扱えるストレージを提供する方法が考えられます。例えば、ストレージのミラーリング技術を発展させ、複数拠点からの同一ストレージデータへのアクセスを可能とすることで、ゲスト計算機に対する変化を抑えるとともに、ストレージの耐障害性の向上も図ることができるはずです。

### 3.7.2 利便性の問題

今回移動通信技術としてNEMO BSを用いましたが、この方法には2つの問題点があります。

まず、NEMO BSを含むMobile IP系の移動通信技術一般に言えることですが、HAを介したトンネル通信が前提となってしまう点です。HAが一点障害になるため、HAの多重化技術を導入しなければならない等、別途考慮すべき点が発生します。ただし、この問題は、トンネルを用いない移動通信技術を採用することで解決できる可能性があります。例えば、HIP[*10]やLIN6[*11]、MAT[*12]等を基礎としたMRを用いる方法です。仮想計算機のオフリンクライブマイグレーションを実現するための条件は、仮想計算機が接続する仮想スイッチのネットワーク環境を維持することです。したがって、それを実現する方法がNEMO BSである必然性はありません。

2つ目の問題点は、NEMO BSを利用したことで、すべてのゲスト計算機が同一の固定ネットワークのノードとして管理される点です。これによって、ホスト計算機が移動した場合、関連するすべてのゲスト計算機群も同時に移動しなければなりません。この問題は、次のいずれかの方法で対応可能と考えられます。1つはゲスト計算機がホスト単位の移動通信技術、例えばMobile IPを採用することです。この場合、すべてのゲスト計算機でMobile IPへの対応が必要となり、導入に関する敷居が高くなります。NEMO BSを利用すると、ゲスト計算機の変更が不要になり、標準のゲスト計算機をそのまま利用できるという大きな利点があります。これらの折衷案としてゲスト計算機ごとにNEMO BS環境を提供することが考えられます。一種のMobile IPプロキシのような運用です。この方法を使うと、ゲスト計算機の変更を避けながら、ゲスト計算機単位での移動ができるようになります。

## 3.8 終わりに

今後到来するクラウド環境において、計算機資源の流動性を確保することが重要です。仮想計算機のライブマイグレーション技術は、資源流動性を確保するための有力な候補です。ただし、現在のライブマイグレーション技術では、移動先が同一セグメント内の別のホスト計算機に限定されています。これは、ゲスト計算機に提供されるネットワーク環境が、ホスト計算機が接続する物理ネットワークに依存するためです。この制限を取り払い、別セグメント、遠隔のセグメントに移動できれば、計算機資源をより流動的に管理できるようになります。ここでは、仮想化技術とNEMO BS技術を用いて、セグメントを共有していない複数のホスト計算機間でゲスト計算機を移動させる手法を提案しました。NEMO BS技術を使い、ホスト計算機の接続ネットワークにかかわらず常に固定のネットワーク環境をゲスト計算機に提供することで、セグメントを越えた移動が可能になります。

最後に、本研究を進めるにあたり、中村雅英氏、Jean Lorchat氏、Martin André氏からLinuxおよびMIPL/NEPLの構成について多くの助言をいただきました。また、実験環境の準備に尽力していただいた三宅喬氏、織学氏およびInterop Tokyo 2009 NOCチームのみなさまに感謝いたします。

執筆者：
**島 慶一（しま けいいち）**
株式会社IIJイノベーションインスティテュート技術研究所
今後ますます進んでいくインターネット端末のワイアレス化に必要となる、IP移動通信技術の研究開発を進めている。

＊10 Robert Moskowitz, Pekka Nikander, Petri Jokela, and Thomas R. Henderson. *Host Identity Protocol*. IETF, April 2008. RFC5201
＊11 Mitsunobu Kunishi, Masahiro Ishiyama, Keisuke Uehara, Hiroshi Esaki, and Fumio Teraoka. LIN6: A New Approach to Mobility Support in IPv6. In *Wireless Personal Multimedia Communication (WPMC)*, November 2000
＊12 相原玲二, 藤田貴大, 前田香織, 野村嘉洋. アドレス変換方式による移動透過インターネットアーキテクチャ（〈特集〉次世代移動通信ネットワークとその応用）. 情報処理学会論文誌, Vol. 43, No. 12, pp. 3889-3897, 20021215