

第 XXIV 部

クラウドコンピューティング基盤の 構築と運用

第 24 部

クラウドコンピューティング基盤の構築と運用

第 1 章 ワーキンググループ設立にあたって

本章では、WIDE cloud ワーキンググループ設立の背景を述べ、ワーキンググループ設立の理由と活動の趣旨に関して述べる。

1.1 ワーキンググループ設立の背景

WIDE cloud ワーキンググループは、2009年10月から実質的な活動が開始された。背景として、近年クラウドコンピューティングと呼ばれる技術が普及し始めており、数々のサービスが登場し始めている。具体的には、Google App Engine[35] や Amazon EC2[7]、salesforce[86] といったものが存在する。これらは無償もしくは有償にて提供されるクラウドサービスであり、特に Google App Engine や Amazon EC2 は、誰でも手軽に利用できるパブリッククラウドサービスの代表例となっている。

Google App Engine は Web アプリケーションに特化した SaaS (Software as a Service) と呼ばれる形態のクラウドサービスであり、Web アプリケーション開発ならびに運用基盤を提供するクラウド網となっている。また、Amazon EC2 は PaaS (Platform as a Service) もしくは IaaS (Infrastructure as a Service) と呼ばれる形態のクラウドサービスであり、サーバ仮想化技術を用いた仮想マシンを提供するクラウド網となっている。これらのサービスは広く一般的に用いられ始めており、商用サービスとして展開され始めている。

これらサービスを構築している基盤技術は、Ajax や仮想化ハイパーバイザー、分散データベースといった既存の技術である。既存の技術をうまく組み合わせることで、ユーザにそのサービスの提供地点を意識させない、網としてのサービス提供を実現することが「クラウドコンピューティング」と呼ばれている。つまり、クラウドコンピューティングは既存の ASP やサーバ仮想化といったものを、より広域に利

用できるようにしたものである。

1.2 ワーキンググループ設立の趣旨

前述の背景をふまえ、さらなる広域化や分散化に対応するための技術を検証、開発するために設立されたのが WIDE cloud ワーキンググループである。WIDE cloud ワーキンググループが目的とするクラウド網とは、WIDE の研究に適したテストベッドとサービスを提供するためのクラウド網であり、そのために必要な要素技術の研究開発、ならびにその実運用を行う。具体的には、サーバ仮想化技術を利用した VM クラウド網を構築し、以下の技術を適用するにあたっての要素技術ならびに運用技術に関する研究開発を行う。

- 広域分散ストレージの開発ならびに実証実験
- 拠点単位の冗長性を有したサーバ、ストレージの運用技術の実証
- ネットワーク透過性を有したサーバマイグレーション
- 拠点間でのリソースの融通と管理
- WIDE の研究開発に適した VM ハイパーバイザー統合管理システムの開発

以上の技術をもとに、図 1.1 に示すような、大学間に広域にまたがった VM クラウド網 (WIDE クラウド) を構築することを目指す。

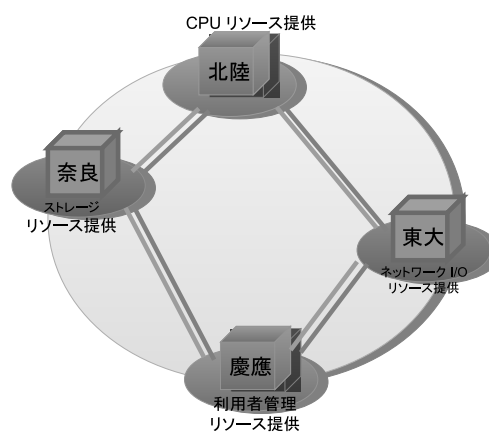


図 1.1. WIDE クラウド概念図

 第 2 章 2009 年活動の概要

本ワーキンググループは、2009 年 10 月からワーキンググループとしての活動を開始した。まず、要素となる技術の調査と、クラウド網の基盤となる仮想化技術の実証から開始した。また同時に、WIDE クラウドのあるべき形を議論し、そのまとめを行った。

2.1 WIDE クラウドに求められる要件

WIDE クラウドは、大学間にてリソースを融通し合い、広域にまたがったひとつの VM クラウド網を形成することを目指す。このクラウド網は、WIDE の日々の研究活動ならびにサービス提供に利用できるものであるべきである。そこで、2009 年 11 月 WIDE 研究会にて BoF を開催し、WIDE クラウドに求められる要件に関する議論を行った。BoF での議論とメーリングリストでの議論の結果、以下の要件があげられた。

- 研究に使える Amazon EC2 のような VM クラウドサービスを提供すること
- 気軽に OS 入りの VM を立ち上げ、実験に使えること
- 拠点単位の停電や障害に影響されないサーバ運用ができること
- 拠点間でのネットワーク透過性を実現すること

2.2 WIDE クラウドの構築

2009 年 12 月から順次機材が納入されたため、各拠点にて WIDE クラウドの構築が開始された。前述の要件定義をうけ、各拠点に iSCSI ストレージと VM ハイパーバイザー用サーバを設置し、VM ハイパーバイザーとして XenServer[21] もしくは Xen[13, 100]、KVM[37, 62] を導入した。

各拠点の iSCSI は、全体を 2 分割し、それぞれを (1) 各拠点の VM ハイパーバイザーからマウントして使うための領域、(2) 各ボリュームを連結して一つの大きなボリュームとして構成し、各拠点の VM ハイパーバイザーからマウントするための領域、として構成した。

現在、各拠点ごとに VM ハイパーバイザーの検証や、iSCSI のパフォーマンス測定を行っている段階で

ある。また、各拠点の VM ハイパーバイザーを統合管理するための、コントローラの検証も開始している。具体的には、Eucalyptus[77, 97] や OpenNebula[27]、virt-manager[2]、oVirt[1] といった管理システムの検証を行い、WIDE クラウドに適用するにあたって不足している機能や改善すべき機能を中心に検証を行った。

その結果、前述の WIDE クラウドに求められる要件を満たすためには、現時点では既存のコントローラ実装では機能が不足しているということがわかった。特に、ネットワークのマネージメントに関する機能に関して、各実装ともに柔軟性に欠けており、自由にネットワーク構成を決定できないことがわかった。WIDE クラウドはネットワークの実験、ならびに実環境のサービスを提供するためのクラウドであり、各 VM 毎に利用するネットワークを柔軟に設定できる必要がある。この要件を満たし、かつライブマイグレーションを行うことのできるコントローラ実装は存在しなかった。

そこで、ワーキンググループとしては、WIDE クラウドに適したコントローラを独自に研究開発する方向で今後の活動を進めていく方針である。また、WIDE クラウドの重要な要素技術となる、ネットワーク移動透過性に関して、重点的に研究開発と検証を行う方針である。

この方針を受け、2009 年は、ネットワーク移動透過性に関する実証実験を中心とした活動を行った。

 第 3 章 NEMO を用いた移動透過性の実証実験

2009 年は、WIDE クラウド構築の要素技術のひとつである、ネットワーク移動透過性に関する実証実験を行った。具体的には、WIDE の研究開発成果のひとつである NEMO[26] を用いて、VM ハイパーバイザーにネットワーク移動透過性の機能をもたせることで、その上で動作する VM にネットワーク移動透過性を持たせることを可能とした。

本章では、NEMO を用いた移動透過性の実証実験に関して述べる。なお、本章にて述べている実証実験は、論文 [112] にて述べられている実験と同様のものである。

3.1 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おり、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る。以後、本章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、稼働中のゲスト計算機を異なるホスト計算機間で移動させる機能をライブマイグレーションと呼ぶ。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ぼ無停止で別の機材に移動できる。ただし、現在提供されている機能は、ホスト計算機の配置に制限があり、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。これは、ゲスト計算機のネットワーク環境が、ホスト計算機のネットワーク環境に依存してしまうためである。

この制限により、別のセグメントに計算資源に余裕のあるホスト計算機が配置されていたとしても、その資源を利用するためにゲスト計算機を移動させることができない。また、ゲスト計算機を、実際の利用者に近いネットワークに移動させて通信遅延を抑えるなどの、資源再配置もできない。

3.2 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限定されるのは、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ある。よって、ゲスト計算機のネットワーク環境がホスト計算機のネットワーク的な位置によらず一定になるように設計すれば、異なるサブネットワークに接続しているホスト計算機間でもゲスト計算機を移動することができるようになる。固定のネットワーク環境を提供するため、本章ではNEMO BSを用いる方法を提案する（図 3.1）。

ホスト計算機は移動ルータ（MR）としての機能を有し、インターネット接続用（a）と、固定アドレス空間提供用（b）のふたつのインターフェースを持つ。インターフェース（b）は、物理的なインターフェースではなく、ホスト計算機が提供するゲスト計算機用仮想ブリッジに接続された仮想インターフェースとして構成する。仮想インターフェースおよび仮想ブリッジに接続されたゲスト計算機には、MR が管理する固定アドレス空間のアドレスが割り当てられる。ゲスト計算機のライブマイグレーション時には、ホスト計算機のMR機能も同時に移動する。NEMO BSの機能により、仮想ブリッジ内のア

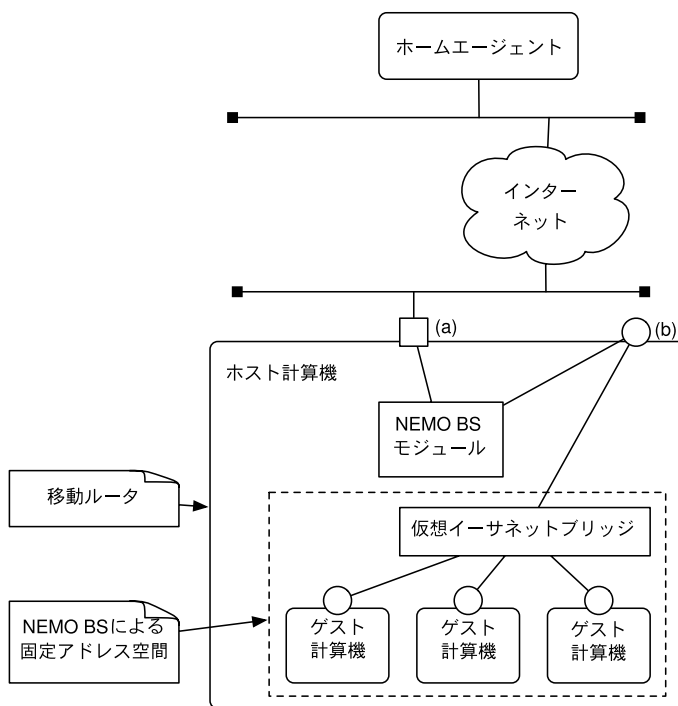
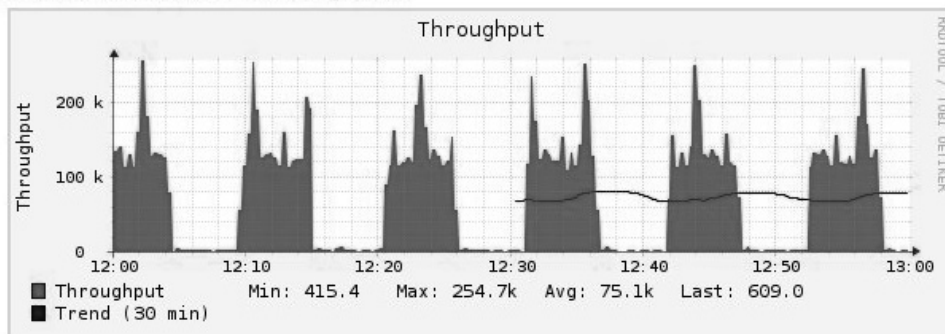


図 3.1. Xen を用いたゲスト計算機オフリンクマイグレーション設計

Xen Server 1: IPv6 UDP streaming traffic



Xen Server 2: IPv6 UDP streaming traffic

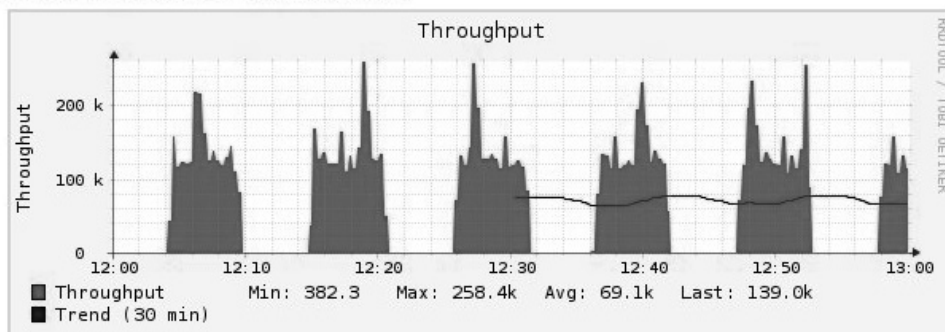


図 3.2. ストリームトラフィックの推移

ドレスが変化しないため、ゲスト計算機がネットワーク環境の変化を認識することなく、継続して稼働できる。

3.3 検証と評価

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実施した。実験環境は、ホームエージェント、NEMO BS 機能を備えた 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の合計 4 台の計算機で構成される。それぞれの計算機はインターネット上の異なるセグメントに接続され、相互に通信が可能な状態となっている。Xen ホスト計算機上には 1 台のゲスト計算機が構成され、ストリーミングサーバとして動作している。

実験では、ゲスト計算機を 5 分毎にライブマイグレーションし、その際のホスト計算機のトラフィックを確認した(図 3.2)。トラフィックグラフが 5 分毎にふたつのホスト計算機の間を移動しているのが確認できる。

3.4 実証実験の考察

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確保が重要な技術となる。ライブマイグレーションを用いて、仮想計算機を別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。本章では、NEMO BS 技術を使い、ホスト計算機の接続ネットワークに関わらず、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マイグレーションを可能にした。提案技術は実際のネットワーク環境を用いて運用され、手法が実現可能であることを示した。

第 4 章 NEMO を用いた実環境サーバのマイグレーション実証実験

さらに、ネットワーク移動透過性の可能性を確認するために、実証実験をもうひとつ行った。これは、WIDE Project の Web サーバである、www.wide.ad.jp サーバをクラウド上に構築し、拠点

間で移動させる実験を行った。

www.wide.ad.jp は現在稼働している実際の Web サーバであり、IPv4 と IPv6 デュアルスタックで全世界に Web ページ公開しているサーバである。

2009 年 12 月 6 日に、www.wide.ad.jp のサーバが設置されている、慶應義塾大学湘南藤沢キャンパスにおいて電源設備法定点検のための停電が計画された。従来であれば、発電機を導入して主要なサーバに対して給電を続け、サーバ運用を行っていた。そこで今回は、WIDE クラウドを暫定的に慶應義塾大学湘南藤沢キャンパスと東京大学情報基盤センターの間にて構築し、停電直前に慶應から東大に対して www.wide.ad.jp VM のライブマイグレーションを行うことで、停電時も特別な対処をすることなく、クラウド上にてサーバ運用を続けるという実証実験を行った。

4.1 クラウド上にてサーバを運用するにあたっての課題

この実験では、以下の 2 つの要求を実現することが技術的な課題であった。

課題 1 実サーバで稼働している OS イメージを、そのまま VM イメージにのせる

課題 2 実サーバが利用している IP アドレスをそのまま利用し続ける

課題 1 を解決するためには、現在実サーバにて稼働している www.wide.ad.jp を VM イメージに変換する必要があった。今回は、MondoRescue[70] というツールを利用した。これは、VM イメージに変換したい OS の上で動作させるツールであり、OS を稼働させたまま VM イメージを作成できるツールである。このツールを利用し、www.wide.ad.jp を VM イメージに変換し、Xen 上にてそのまま動作させることに成功した。

課題 2 を解決するためには、慶應大学、東京大学どちらにも現在の www.wide.ad.jp が存在するネットワークセグメントを提供する必要がある。通常の解決方法としては、VLAN 等を利用し、www.wide.ad.jp が存在するネットワークセグメントを、慶應ならびに東大に設置された Xen サーバの物理インタフェースまで提供することが必要となる。

しかし、通常的环境では、組織間にまたがった VLAN を構築することは難しく、管理者同士の事前の交渉ならびに設定が必要となる。WIDE クラウ

ドは異なった組織をまたがって構築されるクラウド網であり、VM が利用するネットワークが増えるたびに、各組織間に VLAN を設定するのは現実的ではない。そこで、VLAN を使わずにネットワーク移動透過性を実現する手法として、NEMO を選択した。

4.2 実験概要

NEMO を利用することで、www.wide.ad.jp が利用している IPv4 ならびに IPv6 両方のネットワークセグメントを、VLAN 等を設定することなく、慶應と東大それぞれの組織に提供することが可能となった。実験構成図を、図 4.1 に示す。

この構成にて、以下の手順にて実験を行った。

1. 停電直前まで慶應大学にて www.wide.ad.jp を稼働
2. 停電開始前に www.wide.ad.jp を東大にライブマイグレーション
3. 復電後に www.wide.ad.jp を東大から慶應にライブマイグレーション

この間、www.wide.ad.jp の利用者にとって、停電によるサービス中断が発生したり、ライブマイグレーションしたことを気づかれずにサービスを提供できたかどうかを観測した。具体的には、東大から www.wide.ad.jp への HTTP アクセス状況を、smokeping[79] というツールを用いて測定した。この結果を図 4.2 に示す。

図において (1) の部分で慶應大学から東京大学へのライブマイグレーションが行われた。これによってサーバが計測地点から近くに存在するようになったため、ネットワーク遅延が減少している。また、(2) の部分でサービスの中断が発生した。これは、東大の Xen がハイパーバイザーごと再起動してしまったためである。なぜ Xen サーバが異常終了してしまったかの原因ははっきりしていないが、おそらくカーネルパニックが発生して自動的に再起動したものである。これが Xen のバグなのか、NEMO 実装のバグのせいなのかははっきりしていないため、安定したサービスを提供するためには、今後さらなる検証が必要である。(3) の部分にて、法定点検のための停電が終了したため、東京大学から慶應大学へのライブマイグレーションを行った。これによってネットワーク遅延が変化している。

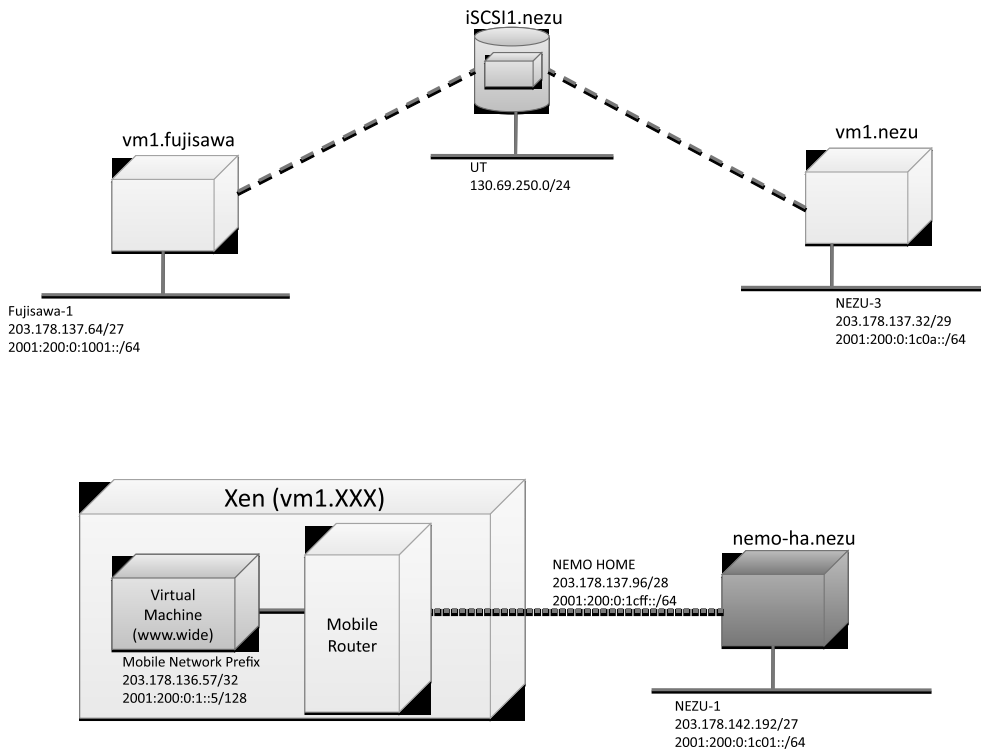


図 4.1. 実環境サーバのマイグレーション実験構成図

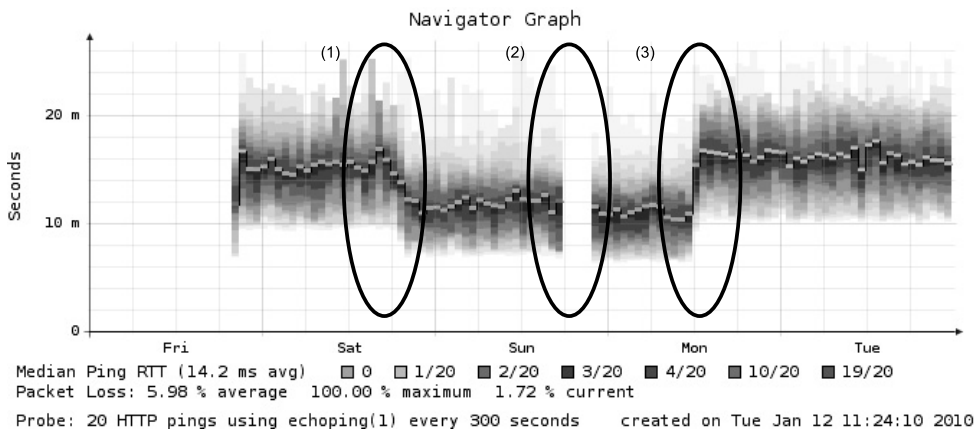


図 4.2. www.wide.ad.jp へのネットワーク遅延計測結果

4.3 実験の考察

今回の実験によって、実サーバにて稼働している OS やサービスを、そのまま WIDE クラウド上に移して運用することが可能であることが確認できた。また、VLAN 等の技術を使わずに、異なった組織間でネットワーク移動透過性を実現することも確認できた。しかし、今回の実験では VM ハイパーバイザーである Xen が異常終了してしまう現象が発生したため、安定したサービスを提供するた

めには、今後のさらなる検証が必要であることがわかった。

第 5 章 まとめ

WIDE cloud ワーキンググループは 2009 年 10 月に発足したばかりのワーキンググループであり、その

活動はまだ始まったばかりである。まずは、WIDEプロジェクトにおける研究とサービスのテストベッドとして利用できるよう、WIDEクラウド網を完成させることをワーキンググループの活動方針とする。

この活動方針に従い、2009年は、まずネットワーク移動透過性技術に関する検証から始めた。この検証と改善をすすめるとともに、コントローラの研究開発に取り組み始めている。2010年も、WIDEクラウドを構築するための基盤技術の研究開発、検証を進めて行き、2010年前半には、WIDEクラウドの試験運用を開始することを目指す。