

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 5 月 31 日現在

機関番号：14401

研究種目：若手研究(B)

研究期間：2010～2011

課題番号：22700052

研究課題名(和文) オーバレイネットワークを応用する広域分散型仮想クラスタ制御機構に関する研究

研究課題名(英文) A study on control structure for wide-area distributed virtual cluster using overlay network

研究代表者

伊達 進 (DATE SUSUMU)

大阪大学・サイバーメディアセンター・准教授

研究者番号：20346175

研究成果の概要(和文)：本研究では、広域環境上に構築される仮想クラスタを構成する仮想計算機を監視・管理しつつ、それら複数台の仮想計算機からなる仮想クラスタ構築技術を、クラスタ管理システム Rocks とオーバレイネットワーク技術 N2N を用いてプロトタイプ実装することに成功した。

研究成果の概要(英文)：This research has succeeded to prototype a virtual-cluster management system that dynamically aggregates multiple virtual machines. The virtual cluster management system is built with Rocks and N2N overlay network technology.

## 交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	1,700,000	510,000	2,210,000
2011 年度	1,400,000	420,000	1,820,000
総計	3,100,000	930,000	4,030,000

研究分野：総合領域

科研費の分科・細目：情報学・計算機システム・ネットワーク

キーワード：グリッド、クラウド、仮想クラスタ構築、仮想計算機、オーバレイ

## 1. 研究開始当初の背景

物理的な計算資源上で複数台の仮想計算機 (Virtual Machine) を稼働させることができる仮想計算機技術が注目を集めている。特に、近年のプロセッサ技術の発展によって、物理的な計算機とその上で稼働する仮想計算機での計算性能差が縮小する傾向にあり、高性能なユーザ占有計算環境を提供できる技術として仮想計算機技術への期待と関心が世界的に高まっている。さらに、このような仮想計算機技術への期待と関心は、クラスタ、グリッド等に代表される分散計算技術との融合により、複数の物理的な計算資源上に仮想的な計算機を配置し、それらを集約する

ことで仮想的な分散計算環境を実現しようとするクラウド技術に関する研究開発を世界規模で活発化させている。Amazon で研究開発が進められている Amazon EC2 (Elastic Compute Cloud)、アルゴンヌ国立研究所やシカゴ大学を中心として研究開発が進められている Nimbus、カリフォルニア大学サンタバーバラで研究開発が進められている Eucalyptus (Elastic Utility Computing Architecture Linking Your Programs To Useful Systems) は、ウェブサービス経由でユーザに仮想的な分散計算環境を提供するクラウド技術の一例としてあげられる。

しかし、これら今日利用可能なクラウド技術は、単一組織内の単一あるいは複数の物理

計算資源上で仮想的な分散計算環境(仮想クラスタ)を動的に構築可能にするものの、複数組織の物理計算資源上の仮想的な計算機を集約・統合した広域分散型仮想クラスタを動的に構築するには課題がある。一般的に、XenやVMWare等の仮想計算機技術を用いて仮想計算機を物理計算機上に構築する場合、当該物理計算機と仮想計算機は物理計算機上のネットワークインタフェースを共有する。そのため、複数の仮想計算機を相互接続しユーザ占有の仮想クラスタ環境として構築する際には、仮想計算機間のネットワークにVPN(Virtual Private Network)やVLAN(Virtual LAN)などの技法を適用しネットワークを保護しなければならない。しかし、今日のクラウド技術を用いて複数の組織にまたがった広域分散型仮想クラスタを構築するためには、計算機やルータ等にVLANやVPNなどの設定を管理者にあらかじめ依頼する必要があり、仮想クラスタの動的な構築を困難にしている。

このような背景から、広域に分散する物理計算資源上に仮想計算機を分散配置し、ユーザの計算環境に対する要求に基づき、これらを動的に集約・統合することでユーザ占有の広域分散仮想クラスタとして構築する手法および技術を模索する潮流が急速にうまれつつある。単一組織内で実現されたクラウド環境を相互に連携しようとするスカイコンピューティング、ユーザの手元の計算資源が突発的に不足した場合にのみクラウド環境で計算を行わせようとするサージコンピューティングは、そのような研究開発潮流の一例としてあげることができる。

## 2. 研究の目的

本申請研究では、広域環境上に仮想計算機資源を分散配置し、オーバーレイネットワーク技術を応用しそれらを動的に集約・統合することでユーザ占有の分散計算環境を提供する仮想クラスタ構築技術の実現を最終目的とし、広域環境上の複数の仮想クラスタを監視・管理・制御する仮想クラスタ制御機構を実現することを目的とする。

## 3. 研究の方法

本申請研究では、すでに開発済みの通信中継機構と連動させ、広域環境上に構築される仮想クラスタの構成仮想計算機を監視・管理し、当該仮想計算機の計算負荷や障害等の状

態変動に応じて、仮想クラスタ間での仮想計算機の動的再配置を可能にする広域分散型仮想クラスタ制御機構(図1)を実現する。具体的には、広域環境上で多数のユーザが仮想クラスタを構築・利用する場合を想定し、仮想クラスタを構成する仮想計算機の障害や計算負荷変動の際に、仮想クラスタ間で仮想計算機を動的に再配置することで、ユーザの大規模分散計算を効率よく無停止で実行可能とする仮想クラスタ構築技術を完成させる。

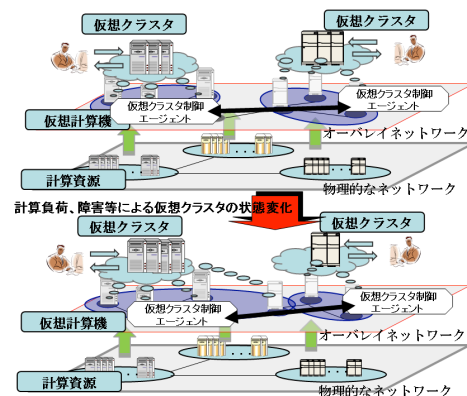


図1 広域分散型仮想クラスタ制御機構

本研究では、上記の目的を達成するため、マイルストーンとなる以下の3つの課題を設定し、研究を遂行した。

[課題1]仮想クラスタ監視、管理、制御のための分散エージェント連携モデルの設計

先行研究で申請者らが開発に成功している仮想クラスタ通信中継機構は、オーバーレイネットワーク技術PIAX(P2P Interactive Agent eXtensions)によって構築したオーバーレイネットワーク上で仮想計算機技術Xenによる仮想計算機間のIP通信を中継する。この通信中継機構は、仮想クラスタを構成する仮想計算機上に配備されるPIAXエージェント(PIAX APIを用いて実装されるプログラム)の相互連携によって実現されている。この通信中継機構との親和性を考慮し、本申請研究で実現する広域型仮想クラスタ制御機構についても、仮想クラスタを構成する仮想計算機内に配置されるPIAXエージェントとして実装するアプローチをとる。本課題では、広域に分散するPIAXエージェントが、どのように仮想計算機上の計算負荷、利用状況を監視し、どのように仮想計算機の情報と交換し、

どのように通信中継機構と連動し仮想クラスタ間での仮想計算機の再配置を効率的に行うのか、という観点から、分散するエージェント間の連携モデルを設計する。

【課題 2】 分散エージェント連携モデルのプロトタイプ実装と動作検証・性能評価

課題 1 で設計した連携モデルのプロトタイプを実装し、広域に分散する仮想クラスタ間で仮想計算機の再配置を行う際の分散エージェント連携モデルの動作検証および性能評価を行う。

【課題 3】 通信中継機構と連携する仮想クラスタ制御機構の実装・評価

本課題では、課題 2 で構築した分散エージェント連携モデルのプロトタイプ実装を基に、すでに開発済みの通信中継機構と連動させた、広域分散型仮想クラスタ制御機構を実装する。また、実装した広域分散型仮想クラスタ制御機構を実際の広域環境上に展開し、実際の科学アプリケーションを応用し、その実用性、有用性について評価を行う。

#### 4. 研究成果

本申請研究では、広域環境上に構築される仮想クラスタを構成する仮想計算機を監視・管理しつつ、それら複数台の仮想計算機からなる仮想クラスタ構築技術を、クラスタ管理システム Rocks とオーバーレイネットワーク技術 N2N を用いてプロトタイプ実装することに成功した。

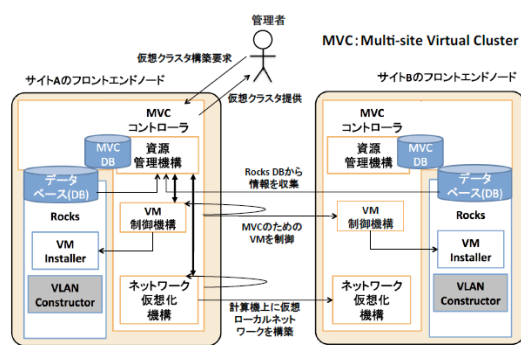


図 2 マルチサイト仮想クラスタ構築システム

本研究でプロトタイプ実装を行った仮想クラスタ構築システムの概要を図 2 に示す。提案・実装を行った仮想クラスタ構築システムは、複数のサイトの計算機から余剰計算資源を仮想計算機としてきりだし、それらをオ

ーバレイネットワークで接続し、シングルシステムイメージとして仮想的なクラスタシステムとして構築する。より技術的には、クラスタ構築ツールである Rocks で構築される複数の物理クラスタ上から、計算負荷の低い計算ノードを選定し、それらノード上で仮想計算機技術 Xen による仮想計算機を稼働させ、それらをオーバーレイネットワーク技術 N2N によって仮想ネットワークで接続し、仮想クラスタとして構築する。

仮想クラスタ構築システムは、資源管理機構、仮想計算機制御機構、ネットワーク仮想化機構の 3 つの分散エージェントから構成される。資源管理機構は、構築される仮想クラスタを構成する計算ノードとなる仮想計算機、および、ネットワーク仮想化機構で用いるオーバーレイネットワークに関する情報を一元的に管理する。本研究では、本研究で基盤技術としたクラスタ管理ツール Rocks が利用する MySQL データベースを拡張し、資源管理機構の情報制御を実現した。ネットワーク仮想化機構は、複数サイトに位置する仮想計算機間でレイヤ 2 ネットワークを構築する。本研究では、ネットワーク仮想化機構で利用するオーバーレイネットワーク技術として、当初オーバーレイネットワーク技術として PIAX を候補技術として考慮していたが、構成されるオーバーレイネットワーク技術の安定性、性能を技術検証した結果、N2N を採用した。仮想計算機制御機構は、資源管理機構が保持するデータベースの情報をもとに、リモートサイトの物理計算機上で仮想計算機を起動、停止するよう実装した。

本研究で提案、実装した仮想クラスタ構築システムで構築した仮想クラスタの性能評価実験を行った。本実験では、1 台のフロントエンドノードと 5 台の VM コンテナからなる Rocks クラスタ 1 基 (A)、1 台のフロントエンドノードと 4 台の VM コンテナからなる Rocks クラスタ 2 基 (B, C) をセットアップした。それぞれのクラスタを構成するノード間は、1Gbps のネットワークスイッチで接続されており、また 3 基の Rocks クラスタはフロントエンドノードが 1Gbps のルータで接続されている。このような環境上で、Rocks クラスタ A のフロントエンドノード上に 1 台の仮想フロントエンドノード、Rocks クラスタを構成する VM コンテナ上でそれぞれ 1 台ずつ仮想計算機を稼働させた仮想クラスタを構築し、分散計算型の科学計算アプリケーションである DOCK を用いて性能評価を行った。性能評価実験では、あるターゲットとなるタンパク質に対して、400 個の化合物のドッキング

シミュレーションを行った際に要した処理時間を計測した。その際、3基のRocksクラスタを接続するネットワーク上で意図的にネットワーク遅延を加え、広域環境をエミュレートしつつ、処理時間の変化を観測した。

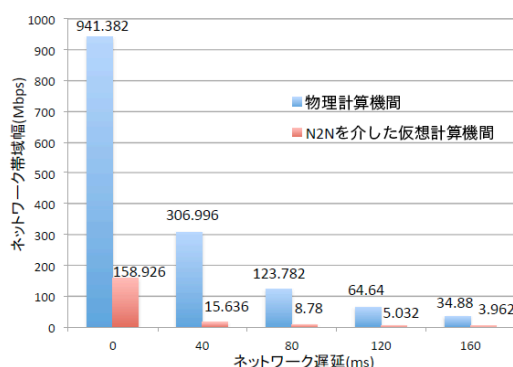


図3 分散計算型アプリケーションを用いた計算性能の評価

実験結果を図3に示す。横軸はネットワーク遅延、縦軸はDOCKの処理時間を示す。この結果から、構築された仮想クラスタは広域環境下においても分散計算型アプリケーションの処理性能には大きく影響しないことを確認した。同時に、これらの計算性能は、実用的であることもまた確認した。

#### 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計5件)

① 多田大輝, 市川晃平, 伊達進, 阿部洋文, 下條真司, “オーバーレイネットワークを用いたマルチサイト仮想クラスタ構築システム”, 先進的計算基盤システムシンポジウム(sacsis2012), 神戸, 2012年5月18日。

② Pongsakorn U-chupala, Kohei Ichikawa, Hirotake Abe, Susumu Date and Shinji Shimojo, “A virtual cluster manager using a hierarchical management model for cloud infrastructure”, The 6th International Conference on Ubiquitous Information Technologies and Applications, pp. 223-228, Seoul, Korea, Dec. 15 2011.

③ Susumu Date, Taiki Tada, Kohei Ichikawa and Shinji Shimojo, “Towards multi-site virtual cluster deployment”, Korea-Japan e-Science and Cloud Symposium,

Proceedings of The 6th International Conference on Ubiquitous Information Technologies & Applications, pp. 29-32, Seoul, Korea, Dec. 15 2011.

④ Kohei Ichikawa, Susumu Date, Yasuyuki Kusumoto, and Shinji Shimojo, “A P2P-based virtual cluster computing using PIAX”, Proceedings of The 2011 IEEE International Conference on Granular Computing, pp. 294-299, Kaohsiung, Taiwan, Nov. 10 2011.

⑤ Wen-Wai Yim, Shu Chien, Yasuyuki Kusumoto, Susumu Date and Jason Haga, “Grid Heterogeneity in In-silico Experiments: An Exploration of Drug Screening Using DOCK on Cloud Environments”, Healthgrid Applications and Core Technologies (Proceedings of HealthGrid 2010), IOS Press, pp. 181-190, Paris, France, June 29 2010.

#### 6. 研究組織

##### (1) 研究代表者

伊達 進 (DATE SUSUMU)

大阪大学・サイバーメディアセンター・准教授

研究者番号：20346175

##### (2) 研究分担者

( )

研究者番号：

##### (3) 連携研究者

市川晃平 (ICHIKAWA KOHEI)

大阪大学・大阪大学情報基盤本部・助教

研究者番号：90511676