課題番号 13-IS05

分散クラウドシステムにおける遠隔連携技術

棟朝 雅晴（北海道大学）

概要　各大学が有するプライベートクラウドシステムを連携させる事で、全国規模での大規模なインタークラウドシステムを実現する事を目的として、クラウド基盤管理ミドルウェア間での連携方式に関する検討、仮想ネットワーク技術などネットワーク連携方式に関する検討、バーチャルマシン群を遠隔接続したバーチャルプライベートクラウドの構築技術に関する検討、大規模分散データベースやストレージシステムの構築技術に関する検討を中心に、必要とされる技術的課題について検討、検証を進めた。本研究の成果として、大学間クラウド連携の実現に向けた要求要件の整理、および、北海道大学情報基盤センターの運用システムの拡充として「ペタバイト級データサイエンス統合クラウドストレージシステム」への展開をはかった。

1. 研究の目的と意義

本共同研究課題は、地理的に分散配置されたプライベートクラウドシステムを連携させることで、大規模なインタークラウドシステムを実現するために必要となる技術的課題について検討、検証することを目的とするものである。学際大規模情報基盤共同利用・共同研究拠点は、スーパーコンピューターや大容量ストレージ、それらを支える計算資源が大容量ネットワークで接続されており、さらに近年、北海道大学アカデミッククラウドをはじめとしていくつかの拠点においてプライベートクラウドシステムの整備が進められており、特に北海道大学におけるクラウドシステムにおいては、コミュニティクラウドとして、全国共同利用により全国の研究者が自由に利用できる環境を整備している。今後、他大学のクラウドシステムとの連携により、さらに大規模な学術インタークラウドシステムを実現するため、以下の技術的課題にについて研究開発を行う。

1. 全国規模でクラウドシステムを相互接続した、大規模なインタークラウドシステムを実現し、運用するために必要とされる技術的課題について検討、検証する。
2. インタークラウドシステムの実現に必要となる、Shibboleth をベースとしたシングルサインオンによる認証連携技術について検討、検証する。特に、クラウド管理ミドルウェアの API レベルでの認証連携技術について検討、検証する。
3. インタークラウド環境下において必要とされる、仮想ネットワーク技術、SINET4 上のオンデマンドVPNなどの活用について検討、検証するとともに、拠点間の相互接続検証実験を実施する。
4. 分散配置され、相互接続されたバーチャルマシン群を用いたバーチャルプライベートクラウドシステム設計法について検討する。特に、大規模分散クラウドシステム上における MapReduce や MPI 等のバーチャルマシンクラスタの構成に関する検討を行い、実験的にクラスタを構成し、その性能について評価を行う。
5. 大規模分散クラウドストレージや分散データベースの実現に必要な、拠点間でのストレージシステムやデータベースの連携技術について検討、検証する。

クラウドコンピューティングにおいて、「規模の経済」は本質的な必要条件であり、規模の拡大などに柔軟に対応できるスケーラビリティは必須である。研究開発においては特に先進性が求められることから、必要に応じてできる限り大規模なクラウド資源を柔軟かつ速やかに利用できることが極めて重要である。したがって、予算上の制約等から各大学において比較的小規模なクラウドシス

テムが独立に運用され、相互利用ができないままでいることは好ましくない。

具体的なユースケースとしては、各拠点に分散配置されているスーパーコンピューターとの連携によるプレ・ポスト処理、ビッグデータ処理など大規模な資源を短期間利用するプロジェクトでの活用、センサー系やネットワーク系の研究など分散型のクラウドシステムが本質的に必要となるプロジェクトで活用などがあげられ、研究プロジェクトの進展に応じて使用する計算資源の規模を柔軟に変更できたり、より好ましい実行環境にシームレスに移行できたり、ある拠点の障害時に自動的に別の拠点に移行ができたり、別の研究プロジェクトで開発されたシステムと連携できたりする必要がある。そのような目的を達成するために、日本全体で連携して運用可能な大規模分散クラウドシステムを実現することが強く求められており、本共同研究における分散プライベートクラウドシステム連携の成果を活用することが期待される。

2. 当拠点公募型共同研究として実施した意義

(1) 共同研究を実施した拠点名および役割分担

北海道大学：北海道大学アカデミッククラウドシステムにおける資源提供、およびクラウドコンピューティング基盤技術に関する支援
東京大学・東京工業大学・大阪大学・九州大学：遠隔クラウド連携に関する支援

(2) 共同研究分野

超大規模情報システム関連研究分野

(3) 当公募型共同研究ならではという事項など

本共同研究課題は、ネットワーク型拠点の特徴を最大限活用し、それぞれ地理的に分散配置されたシステムを活用し、大規模なインタークラウド試験システムを構築するところにその特徴と意義が存在する。大学・研究機関においてプライベートクラウドシステムの構築が進められているところであるが、予算上の制約等により、それぞれのシステムの規模は比較的小規模なものに留まることが多く、クラウドコンピューティングによるスケールメリットが生かされないという問題が生じるが、本研究課題においては、各大学においてそれぞれ導入されたプライベートクラウドシステムを連携させ、大規模なアカデミッククラウドシステムを構築することで、各大学のポリシーを生かしつつ、かつスケールメリットを享受できるような分散環境を実現することを目標としている。そのために必要な技術的課題について、具体的な広域分散クラウドシステムの連携を通して検討、検証するところに、本共同研究の意義が存在する。

3. 研究成果の詳細と当初計画の達成状況

(1) 研究成果の詳細について

(a) インタークラウドシステムの認証連携に関する検討、検証

地理的に分散配置された大学クラウド間の連携によるインタークラウドシステム上でシングルサインオンを実現するための認証基盤に要求される要件を整理し、Shibboleth(IdP, SP)および代理証明書リポジトリ、クラウド管理システム、インタークラウドポータルから構成される認証基盤のアーキテクチャについて検討、検証を行った。(図 1 を参照)

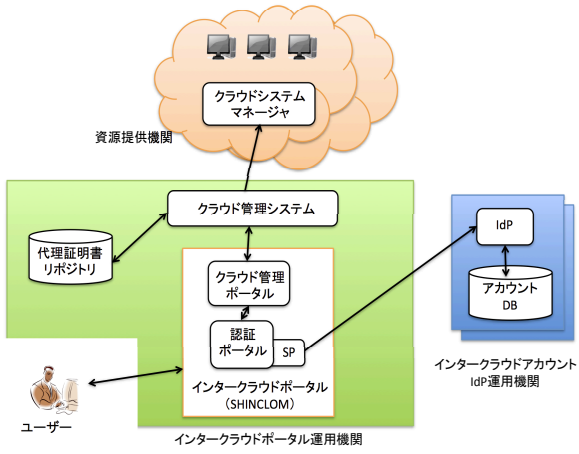

図１ 認証基盤のアーキテクチャ（研究成果(4)[1]より）

RightScale(http://www.rightscale.com/) や Scalr (http://scalr.com/)など現状のマルチクラウドコントローラにおいては、プライベートクラウドシステムマネージャやパブリッククラウドサービスから手動で認証鍵を取得し、それをそのままポータルシステムから登録するという手続きが行われており、ユーザの手間という観点のみならず、セキュリティ上の問題を生じる可能性がある。

Shibboleth IdP, SP を用いる事で、シングルサインオンによる利便性の向上が期待されるが、クラウドシステムに対する API アクセスに必要となる鍵情報について別途管理する必要が生じる。そこで、グリッドコンピューティングで用いられている代理証明書の考えを用い、その有効期限を設定することで、より厳しいセキュリティ要件を満足するシステムの構築、設定、管理を可能としている。代理証明書リポジトリとしては、MyProxy を用い、Shibboleth およびインタークラウドポータルとの連携部分についてシステム構築を行った。

図２にその概要を示す。代理証明書レポジトリと連携して動作する Python のブリッドおよびインターフェイスの部分を自主開発した。さらに、それぞれのクラウド管理ミドルウェアに対応するドライバを作成することで、異なるクラウド環境へ対応した代理証明書の運用管理が可能となる。

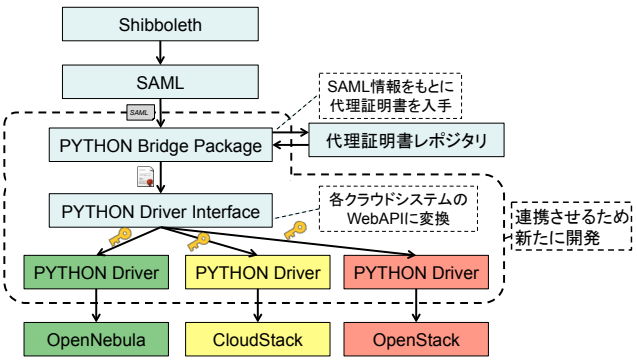

図２ 代理証明書レポジトリを用いた認証連携システムの概要（研究成果(4)[1]より）

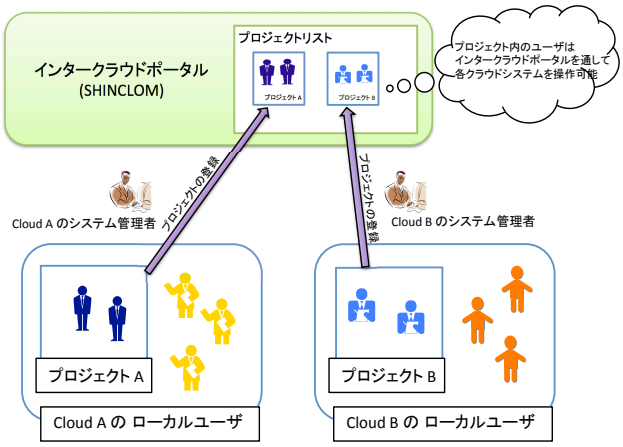

図３ プロジェクト単位でのユーザ管理（研究成果(4)[1]より）

さらに、図３に示す通り、インタークラウドポータル上で、プロジェクト毎のユーザ管理を行う事を可能とする。プロジェクトの考えは、最新の CloudStack (http://cloudstack.apache.org/)においても導入されているが、複数のクラウドシステムを連携させたインタークラウドポータルにおいては、必須の機能であり、グリッドコンピューティングにおける Virtual Organization (VO)に似た機能を実現する事をその最終目的としている。

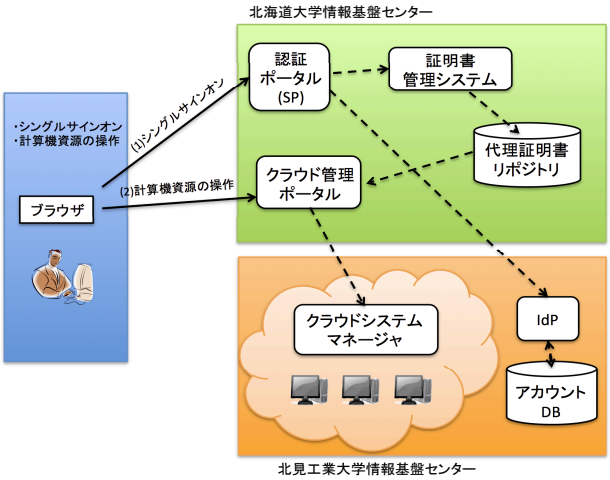

図４ 遠隔実証実験環境の構築（研究成果(4)[1]より）

認証基盤に関する実証実験としては、北海道大

学情報基盤センターと北見工業大学の間で図４に示される遠隔実験環境を構築し、連係動作が正しく行える事を確認している。

(b) バーチャルプライベートクラウド構築に関する検討、検証

ユーザが占有して利用できるバーチャルマシンおよび仮想ネットワークから構成されるバーチャルプライベートクラウド(Virtual Private Cloud, VPC) の構築について検討、検証を行い、図４二示される階層型のアーキテクチャを提案した。本アーキテクチャは、アプリケーションシステムを管理する CASL (Cloud Application Specific Layer), 仮想ネットワークを管理する ADNL (Application Defined Network Layer), 仮想マシンクラスタなど仮想インフラを管理する VOIL (Virtual Overlay Infrastructure Layer), クラウド管理ミドルウェアなどのプラットフォームを管理する CPSL (Cloud Platform Specific Layer)の４層から構成されており、その配下にあるクラウド管理ミドルウェアやパブリッククラウドサービスの違いを吸収した統一的なフレームワークを提示する事で、互いに異なる環境間の連携を実現するものである。

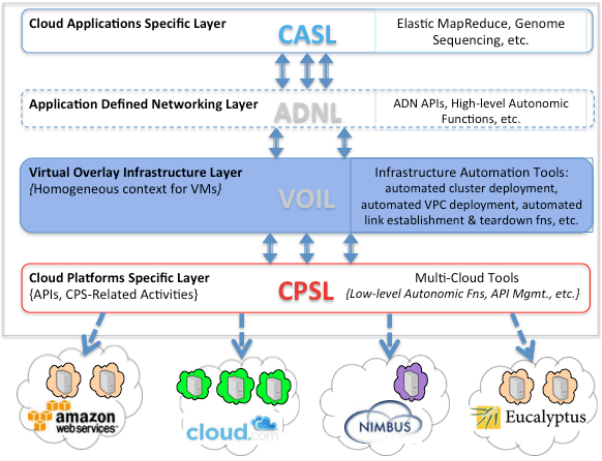

図５ マルチクラウド管理のための階層型アーキテクチャの提案 (研究成果(2)[1]より)

図６に、本アーキテクチャにおいて実現されるバーチャルプライベートクラウドの構築例を示す。一般にパブリッククラウドサービスにおいては、クラス A のプライベート IP、小規模なプライベートクラウドにおいては、クラス C のプライベート IP がそれぞれのバーチャルマシンに割当られているため、バーチャルプライベートクラウドを相互接続する仮想ネットワークにおいては、クラス B のプライベートアドレスを割り当てる事で、IP アドレスの干渉を起こさずに仮想マシンクラスタのネットワーク接続を実現している事例となる。

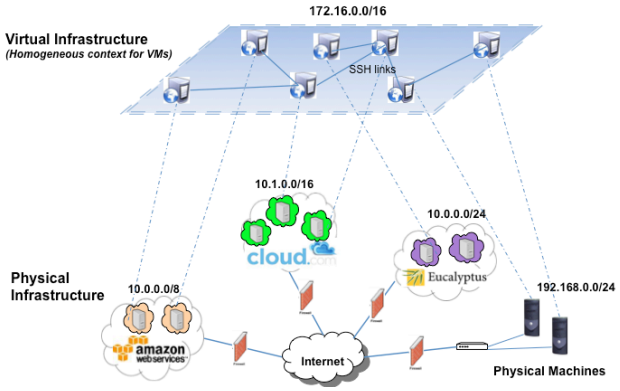

図６ バーチャルプライベートクラウドの構築例 (研究成果(2)[1]より)

さらに、インタークラウド基盤上におけるバーチャルマシンなどの資源割当最適化に関する、検討、検証も引き続き行っている。(研究成果(4)[2])

(c) 分散ストレージ・データベースに関する検討、検証

分散ストレージ・データベースに関する検討、検証については、Apache Cassandra (http://cassandra.apache.org/)のノードを全国規模で地理的に分散配置した場合の検討、検証実験を実施した。

具体的には、図７に示されるように、北大、北見工大、琉球大の間を仮想ネットワークとして Vyatta(http://www.vyatta.org/) を用い、OpenVPN (http://openvpn.net/)を使用した暗号化通信により、そのインスタンス同士を Site-to-Site モードで接続した。拠点間の通信遅延は、北大～北見工大間で 10ms、北大～琉球大間

で 51ms、北見工大～琉球大間で 61ms あったが、そのような通信遅延が存在する分散環境かにおいて Cassandra ノードを動作させることに成功した。地理的にデータを分散配置し、複製させる事で、各サイトでのデータアクセス応答時間を削減する（図８）とともに、大規模な災害や障害発生時においてもデータを保全するためのシステム基盤を実現でき、大規模な災害時を想定した検証実験により、その有効性を確かめることができた。

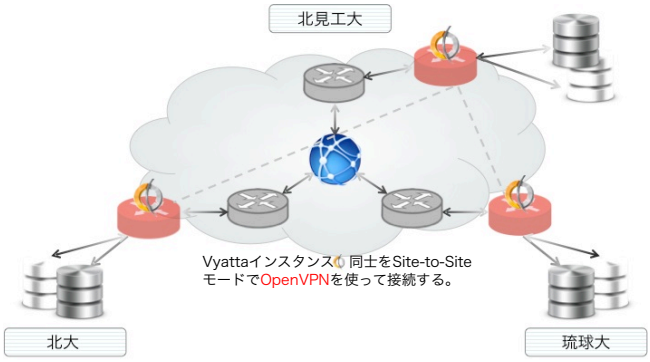

図７　北大・北見工大・琉球大の相互接続による分散 Cassandra クラスタの構成例

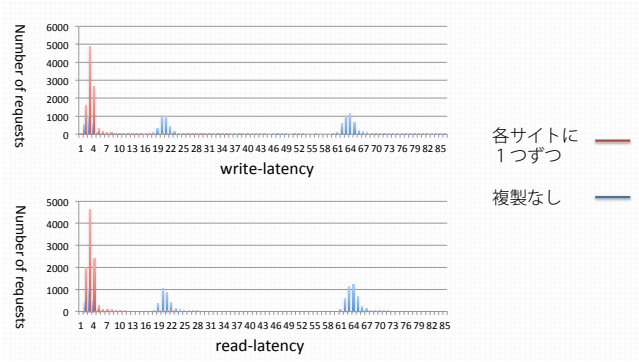

図８　書き込み、読み込みの応答時間（横軸）とその度数（縦軸）の検証結果

さらに、(5)に示す運用システムへの展開に関連して、Amazon S3 互換のオブジェクトストレージの実現について検討をすすめた。

(d) PaaS に関する検討、検証

PaaS (Platform as a Service)に関する検討、検証については、Open PaaS として利用が進められている CloudFoundry の実行環境を北大クラウドシステム上に構築し、ユーザとのインタラクションにより学習をすすめる機械学習アルゴリズムである iGA (interactive Genetic Algorithm)を実装した。さらに、分散キーバリューストアである Redis(http://redio.io/)と連携することで、世界規模の多数のユーザによる利用を想定したスケーラブルなシステムを構築し、最適解の探索、実行を大規模分散環境において処理するシステム基盤を実現した。(図９上を参照)

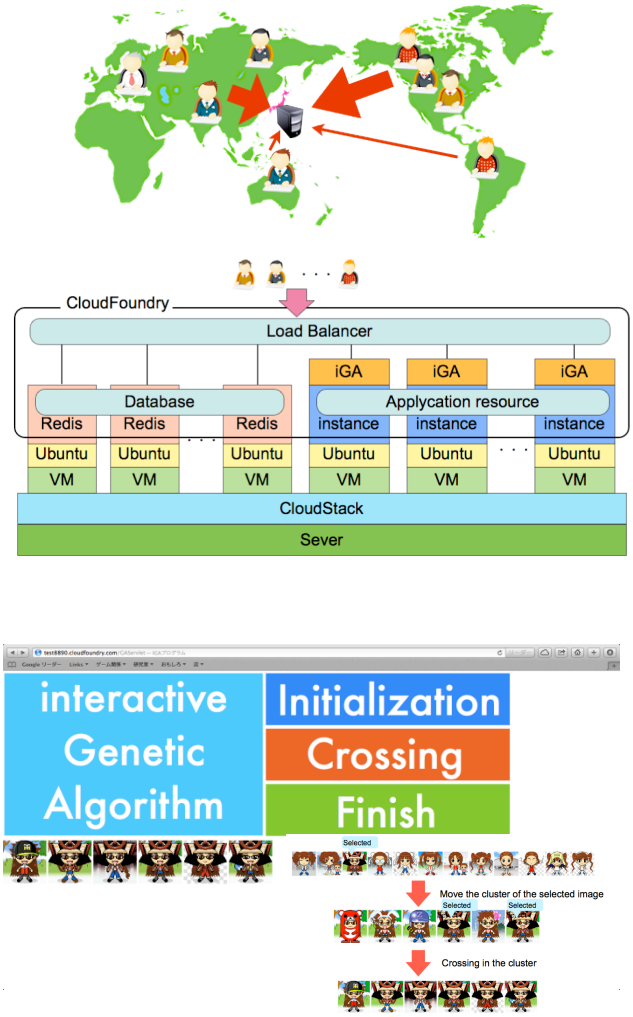

図９ Open PaaS (CloudFoundry)を活用した、世界規模での利用者からのアクセスを可能するインタラクティブ機械学習（iGA）フレームワークの概要（上）とその応用例としてのアバターの自動学習（下）（研究成果(1)[1]より）

iGA は、人間の感性による評価を行い学習するアプリケーション全般に利用できるが、その実例

として、アバターの評価および学習進化を自動化するシステムを構築した（図９下）

今回の実装では、利用可能資源の制約上、北大クラウド内で実行したが、バーチャルマシンクラスタの構成技術と組み合わせる事で、広域分散環境下での実装も可能である。

(e) 運用システムへの展開：ペタバイト級データサイエンス統合クラウドストレージシステムの設計・構築

これまでに得られた知見を参考に、北海道大学情報基盤センターにおける運用システムとして「ペタバイト級データサイエンス統合クラウドストレージシステム」の設計、構築作業をすすめた。

本システムは、図１０に示すように、ペタバイト級のオブジェクトストレージに加えて、数百コア規模のクラウド IaaS 基盤システムから構成され、遠隔からの API アクセスによるストレージ及びクラウド資源の制御を可能とした。

本システムにおいては、Shibboleth による認証連携に対応するとともに、遠隔からの Web サービス API によるアクセスによる連携利用、および Amazon S3 互換 API によるオブジェクトストレージとして遠隔サイト間でのデータ連携が可能となっており、JHPCN を中心とした分散クラウド環境の連携による共同研究支援環境を整備することができた。

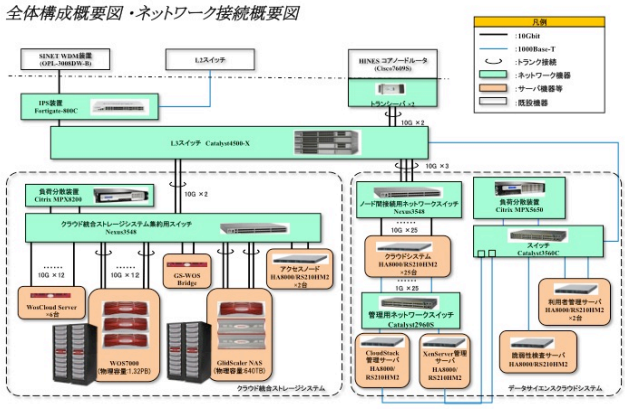

図１０　ペタバイト級データサイエンス統合クラウドストレージシステムの概要

本システムで提供されるサービスについては、他に gfarm による分散ファイルシステム、Hadoop, Hive, Hbase, Mahout, R などをプレインストールした大規模並列分散データ処理システムサービスなどがあげられ、いわゆるビッグデータの関する研究を支援する環境を整備し、HPCI/JHPCN 向けに２０１４年４月より提供しており、クラウド関連の媒体でも紹介されている（研究成果(5)[1]）。

(2) 当初計画の達成状況について

これまでの進捗状況としては、中間報告の時点でほぼ当初の目標を順調に達成し、分散クラウドシステムの連携に必要となる各種技術について検討、検証を進める事ができた。

さらに、中間報告以降においては、ペタバイト級データサイエンス統合クラウドストレージシステムの設計及び構築へのノウハウの展開を中心に作業をすすめることで、北海道大学情報基盤センターにおける運用システムにその成果の一部が反映されており、今後の共同研究の支援も含めた波及効果が期待されている。

4. 今後の展望

本研究課題としては、当初の目的を達成しつつあるものと判断し、本年度で完了した。今後の展望としては、全国規模のアカデミッククラウドインフラ整備への展開、および国際連携も含めたクラウド連携の促進があげられる。

5. 研究成果リスト

(1) 学術論文（投稿中のものは「投稿中」と明記）

[1] 谷沢智史, 西村一彦, 長久勝, 横山重俊, 吉岡信和: プライベートプラウド監視ツールの開発に向けた一考察, サービスコンピューティング研究専門委員会 第 5 回研究会, 電子情報通信学会, 信学技報, vol. 113, no. 86, SC2013-8, pp. 41-46 (2013)

(2) 国際会議プロシーディングス

[1] Courtney Powell, Masaharu Munetomo, Attia Wahib, and Takashi Aizawa: Constructing a Robust Services-oriented Inter-cloud Portal Based on an Autonomic Model and FOSS, Proceedings of the Workshop on Distributed Cloud Computing (DCC2013) (2013) (accepted)

[2] Courtney Powell, Masaharu Munetomo, Takashi Aizawa: Towards a User Deployable Service-oriented Autonomic Multi-cloud Overlay Infrastructure for Sky Computing, Proceedings of the 2013 International Conference on Grid & Cloud Computing and Applications (2013)

[3] Masataka Mizukoshi, Shitaro Bando, Martin Schlueter, Masaharu Munetomo: Implementation of Multiple Classifier System on MapReduce Framework for Intrusion Detection, Proceedings of the 2013 International Conference on Parallel and Distributed Processing Techniques and Applications (2013)

(3) 国際会議発表

[1] Masaharu Munetomo, Shintaro Bando: A Scalable Infrastructure of Interactive Evolutionary Computation to Evolve Services Online with Data, IEEE BigData 2013, Santa Clara, USA (2013)

(4) 国内会議発表

[1] 相澤 孝至, 棟朝 雅晴: Shibboleth を用いたインタークラウドシステムのための 認証基盤の設計, 第６回インターネットと運用技術シンポジウム, 広島大学 (2013)

[2] 川勝 崇史, 棟朝 雅晴: 分散クラウド環境における SLA を考慮した WEB システムの多目的資源割当最適化, 第９６回情報処理学会数理モデル化と問題解決研究会, 東京工業大学 (2013)

(5) その他（特許, プレス発表, 著書等）

[1] "北大、学術クラウド基盤を構築、PB 級ストレージを全国の研究者へ", クラウド Watch http://cloud.watch.impress.co.jp/docs/news/20140403_642616.html, インプレス 他多数