

12-MD07

大規模計算機空気冷却風速場の実時間解析 と移動型ネットワークセンサー連携による計測融合オペレーション

東田 学 (大阪大学)

概要 これまで、大規模計算機空気冷却風速場の流体シミュレーションにおいて、粗い精度であれば、京 1 ラックぐらいの計算能力で実時間シミュレーションが実現可能であることを確認した。本年度は、絶えず変化する環境下で適応的に重要な計測データ取得ポイントを認識し、そのポイントのデータを取得できるロボット計測システム技術を研究・開発し、計測融合オペレーションの実現を目指すこととしており、そのための要素技術開発に目処を付けた。

1. 研究の目的と意義

これまで本グループでは、格子ガス法超並列計算手法を利用することにより、大規模計算機の空気冷却風速場に関する実時間シミュレーションを、現実には十分確保可能な規模の計算機能力で実現可能であることを示した。さらに、風速計等のセンサーから得られる時系列計測データをもとに格子ガス法計算モデルを時々刻々補正することによって、現実に近いシミュレーション結果を得て、これを実際の空調機等の操作に適応的に反映させる計測融合オペレーションシステムの検討を行ってきた。

本研究の目的は、より精度の高い計測融合オペレーションを実現するために、シミュレーション結果と計測データを大規模分散ファイルシステムに蓄積し、時系列および局所位置情報を踏まえた解析処理を行い、より適切な計算モデル補正を行う手法の研究開発と、より適切な観測データを取得するためのネットワーク制御ロボットによる巡回監視手法の研究開発を行うことである。

具体的には、現実の大規模計算機空気冷却システム室（例：“京コンピュータ”施設の計算機室及び空調機室と大阪大学サイバーメディアセンターの主機室）を解析対象にして、空気冷却風速場の実時間シミュレーション結果を大規模分散ファイルシステムである HDFS に蓄積し並行データ解析を可能とする。一方、簡易気象計測装置を搭載したネットワーク制御ロボットを現実の大規模計

算機室に巡回させ、その巡回観測結果を同様に HDFS に蓄積する。双方のデータを時系列および局所位置情報を踏まえて対比するために、Hadoop によって並行データ解析する手法を研究開発する。これによって、(1) 格子ガス法計算モデルを時々刻々補正することによって、センサーの設置されていない場所も含めた全体の風速場について、現実に近いシミュレーション結果を得ることを目指す。さらに、(2) より精度の高いシミュレーション結果を得るために必要となる計測点を導出する手法を確立し、その地点へネットワーク制御ロボットを派遣することによって、より適切な計測データの取得を行うための制御手法を研究開発する。

以上の成果は、ペタフロップスを超える大規模計算機の空気冷却に関する計測融合オペレーションシステムの構築に資するものであり、これによって、本研究は、(1) 筐体内の電子部品にやさしい過渡変化を実現して高信頼性・長寿命化を目指す運転と、(2) 冷却・空調・電源系を含めた計算機システムプラント全体のエネルギー効率向上を目指す運転の両立に貢献できるものになる。また、(3) 計算機センターの異常時にもネットワーク制御ロボットの派遣による一次オペレーション対応と過渡変化シミュレーションに基づく異常解析が可能になることを目指す。

本研究の意義は、以下のとおりである。

ペタフロップスを超える超並列高速計算機システムでは、通常数十メガワット以上の巨大な発熱

を伴うばかりでなく、その発熱分布は広範囲に広がり、計算機システムのジョブスケジューラによる計算ノードの割当て・解除、各種プログラムのエラー発生、計算機ハードウェアの故障のほか、これらに伴う筐体ファン自動オンオフや、故障の修理保守後の起動、さらには、計算機システムの冷却・空調・電源系におけるトラブル発生で予測できない形状変化をする。他方、従来型の冷却制御は、液冷システム、空冷システムともに、計算機冷却場の入口温度と入口流量を確保し、各筐体では、温度や湿度がある制限値を超えた場合に電源オフにすることで計算ボード等の電子装置の保護を行っている。

今後の大規模計算機システムにおいては、上述のとおり、異常な過渡変化の過程が多様化するため、従来の制御方式では、定格温度や定格湿度から大きなオーバーシュートやアンダーシュートが生じる可能性が高い。これらは、制限値に達すれば、稼働率の低下、達しない場合でも、電子部品の信頼性や寿命の低下を招く。このため、各種センサー情報とシミュレーションから過渡変化の全体像を的確に把握し、冷却システムの制御を適応的に実行する“計測融合オペレーション”を可能にする技術の構築がきわめて重要である。

本実現に際して、これまで環境監視に用いられてきた定点センサーやサーモグラフィによる遠隔センサーに代わり、ホットスポットに環境監視装置を派遣できるネットワーク制御ロボットによる巡回計測手法を導入することは、研究開発が立ち上がりつつあるクラウド・ロボティクス分野との活発な技術交流が期待できる。クラウド・ロボティクスは、多数のロボットがクラウドで情報を共有することによって、自律的な群体制御を行ったり、人間と情報共有するために、Google 社などを中心に研究開発が進められている新領域である。さらに、情報基盤センター群の計算機資源を、クラウド技術を用いたデータ解析に利用するという新機軸へ誘導できる。

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大学名と研究体制

大阪大学、東北大学、九州大学

(2) 共同研究分野

「超大規模数値計算系応用分野」、「超大規模データ処理系応用分野」

(3) 当公募型共同研究ならではの事項など

拠点が関心をもつ計算機室を実験計測の現場にすることによって、“数値シミュレーション”と“実験計測”を同一の者が直接行えるというメリットがある。これによって、計測情報のフィードバックによってシミュレーションの精度をあげつつ、シミュレーション結果に応じて、より効果的な計測ポイントを適応的に見つけ出す“計測融合オペレーション”の効率的な開発を可能にしている。さらに、拠点に導入された新しい資源と新しいサービスの立ち上げを促進するという側面も期待されている。

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

(1-1) 大規模計算機室の環境シミュレーション

本プロジェクトでは、これまで、理研 AICS の「京」計算機室を対象として計測融合オペレーションの実現性を検討してきた。本年度からは、その取組から得られた知見を活かし、大阪大学サイバーメディアセンターにて 2014 年春に竣工を予定している「IT コア棟（仮称）」を実施対象に切り替えて、より実践的な計測融合オペレーションの実現を目指すこととした。

IT コア棟は、総フロア面積 2,000m³ のスーパーコンピュータ運用に特化した独立棟二層構造のデータセンターである。まず今後十年を想定して、消費電力 2MW から 3MW 規模のスーパーコンピュータを運用可能な施設を目指し、図 1 に示すような概念設計を行い、特に、以下のような仕様を盛り込んでいる。

まず、計算機設置エリア（第二層）と電源・空調設備設置エリア（第一層）を完全に分離した二

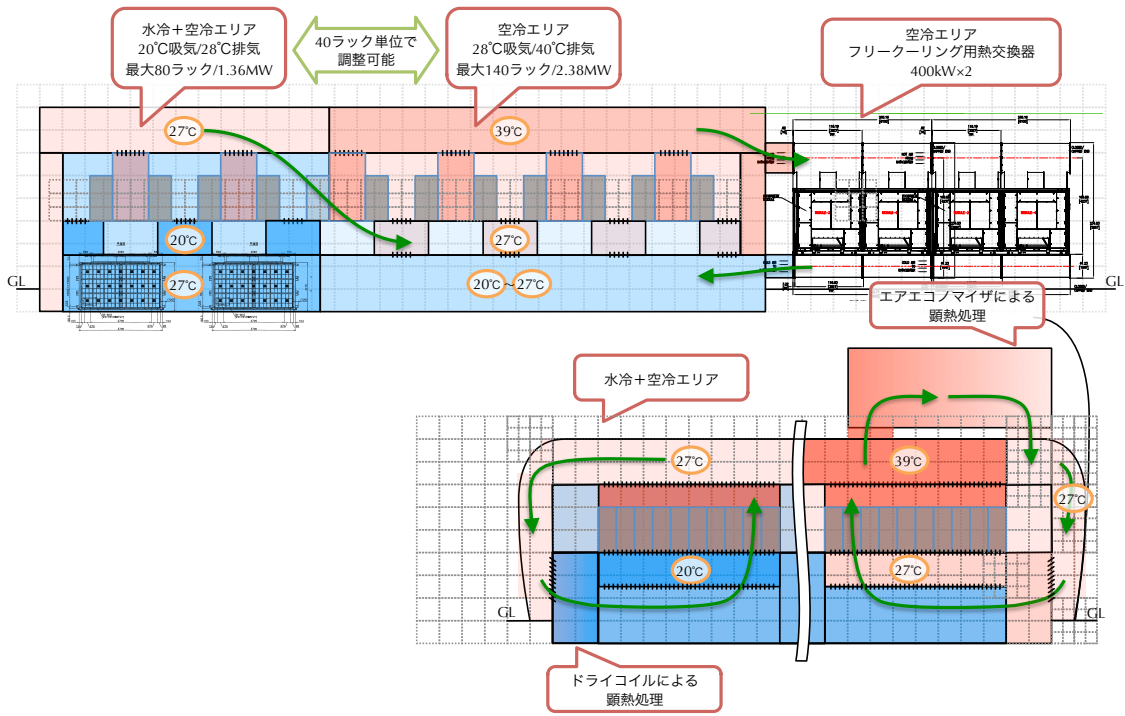


図 1: 阪大 CMC の IT コア棟の概念設計

層構造を取り入れている。これにより、これまでのフリーアクセス・フロア型の計算機室では難しかった配線・配管の適宜的な最適化が可能になると考えている。さらに、空気搬送ロスを最小化するために、自然対流にも配慮した流路を作れる二重天井、二重壁構造を取り入れている。これらの区画を配線・配管用の区画とは完全に独立させることで、最大 5,000m³/min の空気循環が可能になると見積もっている。同時に、排熱を完全に独立して空調設備へ戻すための退路を天井および壁面に設けることにより、高効率の循環を行うことを検討している。さらに、空気流を調整しラック列毎に環境温度を最適化するための配慮や、エアエコノマイザによるフリークーリングにも対応可能な配慮を行っている。

竣工当初には、最大 1MW の空冷負荷を想定しているが、まず、リファレンス・モデル（図 2）の作成を行ない、上述の要求仕様を満たすよう空気による熱搬送路の最適化を行った。モデリングに際しては、竣工時期から逆算される設計期間の制約も考慮し、CFD ソルバ機能も有する Future Facilities 社の 6SigmaDC を導入し熱搬送経路や圧力分布の解析を行っている（図 3、図 4）。

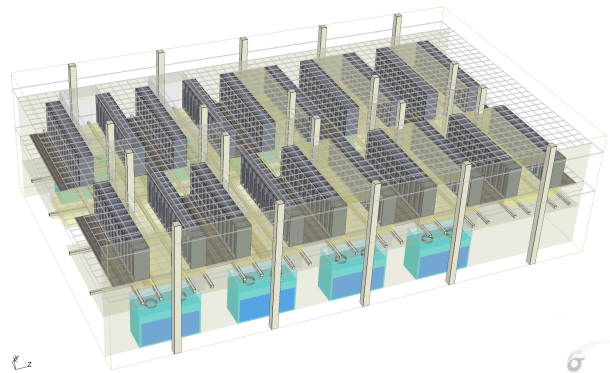


図 2: IT コア棟のリファレンス・モデル

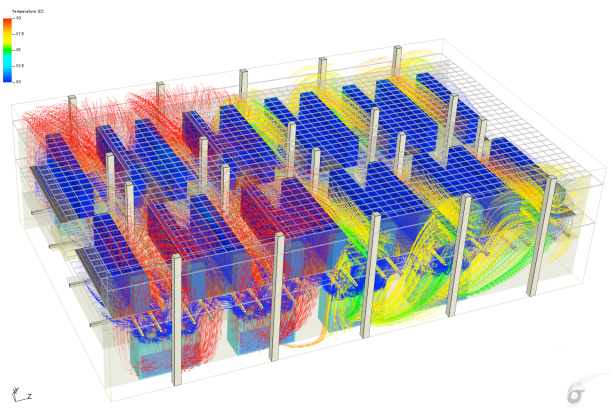


図 3: 熱流線の解析結果

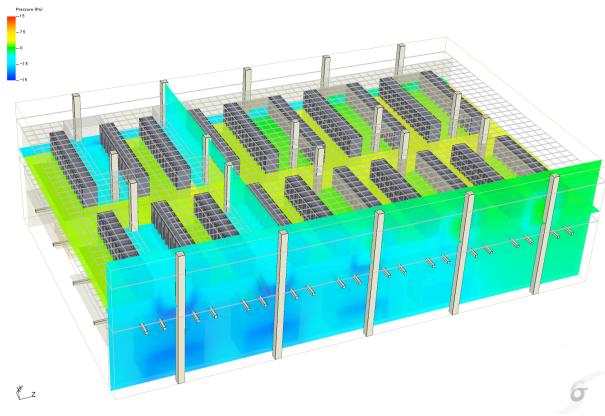


図 4: 圧力分布の解析結果

(1-2) 適応的クラウドロボット技術の開発

ネットワーク制御ロボットは、インターネットに接続された複数のロボットが連携するだけでなく、インターネット上のサービスとも連携することによって、より能動的な情報処理を促進するための枠組みである。

本グループでは、ロボット制御ミドルウェアとして、スタンフォード大学 AI 研究所で開発され、現在は Willow Garage 社が継続的に開発を行っている ROS (Robot Operating System - <http://www.ros.org/>) を研究開発の基盤としている。ROS は、ロボットを構成する様々な機能を提供するノード群が非同期相互通信を行うために XML-RPC による出版・購読 (Publish/Subscribe) 型のメッセージ交換モデルを採用している。これにより、機能拡張に際してノード間の依存性が粗密となっていることもあいまって、オープンソース型プロジェクトとして世界規模の研究開発が行われている。また、ROS は、様々なロボットで動作するが、リファレンスプラットフォームである PR2 の導入には数千万円規模の予算が必要であった。しかし、昨年から Turtlebot という十万円前後で導入可能で、かつ、ハードウェア仕様を開示した普及型ロボットが商品化されており、多数のロボットによる協調作業を検証することが可能になっている。

これらのオープンソフトウェアとオープンハードウェアを基盤として、本グループでは、大規模計算機が設置された計算機室を対象として、絶え

ず変化する環境下で適応的に重要な計測データ取得ポイントを認識し、そのポイントのデータを取得できるロボット計測システム技術を研究開発することを目的としている。より具体的には、ロボットを自律巡回させ経時的な環境情報を取得しシミュレーション結果との整合性を検証する技術の研究開発とより精度の高いシミュレーションを行うためにロボットを派遣し計算モデルを補正する技術の研究開発を行っており、それらを実現するために、次のような要素技術を研究開発している:

- 1) ロボットを自律巡回させ経時的に環境情報を取得する技術
- 2) 計算機室という電磁波伝播の複雑な環境におけるロボットとの無線通信に際して、短期的および長期的な通信途絶に対応可能な通信技術
- 3) ロボットから送出される経時的な情報を蓄積するためのデータベース技術と蓄積された経時データを時々刻々処理し、格子ガス法計算モデルを補正し実時間シミュレーションを繰り返すことによって計算機室全体の風速場について、現実に近いシミュレーション結果を得るための技術
- 4) より精度の高いシミュレーション結果を得るために必要となる計測点を導出しロボットを派遣する技術

以下、各々の要素技術について詳述する。

(1-2-1) 自律巡回

本グループでは、現在、二台の Turtlebot により基礎データの取得と検証を行っている。Turtlebot は、iRobot 社の Roomba Create を台車として、制御用コンピュータとして Linux が稼動するネットブックコンピュータと、視覚センサーとして Microsoft 社の Kinect、また、補助基盤にジャイロセンサーを搭載している。ROS には、Kinect センサーから取得できる 3 次元深度マップ情報を合成処理し、自身の移動に伴って周辺マップを自動的に生成する gmapping 機能が提供されている。さらに、生成された周辺マップを元に、

センサー情報から自身の姿勢を推定する AMCL (Adaptive Monte Carlo Localization) 機能や障害物を回避しながら指定位置へ移動するための遠隔操作機能も実装されており、これらの機能を活用し、自律巡回のための必要機能の検証を進めている。

初代の Turtlebot には、充電ステーションへの自律帰還機能が欠落しているため、行動可能時間が数十分に制限されている。また、走行距離センサーおよびジャイロセンサーの精度が十分ではなく、移動に伴う累積誤差が自己位置推定機能が補正可能な範囲を超えてしまっている。これらの要員により、Turtlebot によって長時間のオペレーションを行うには、随伴要員が不可欠であった。本グループでは、これらの問題を解消するために、独自仕様の Turtlebot の構成検討を進めて来たが、今秋、Willow Garage 社から次世代 Turtlebot の設計仕様が公開され、これらの問題が改修される目途が立ち、年初にも商品化される第二世代 Turtlebot の導入を行うこととした。これにより、より長時間の自律巡回とより精度の高い派遣制御が可能になることを期待している。



図 5: Turtlebot による移動型センサー

(2-2) 通信途絶耐性

ROS で制御されるロボットとの情報交換には、通常、Wi-Fi による無線通信を行う。計算機室のような電磁波伝播の複雑な環境では、安定的なデータ到達性を期待することは難しいことが予想されるため、大阪大学サイバーメディアセンター本館主機室における Wi-Fi 到達性と IP 到達性の検証を進める一方で、無線通信のための電波が途絶した状態を想定したロボット通信制御方式の検討を進めている。

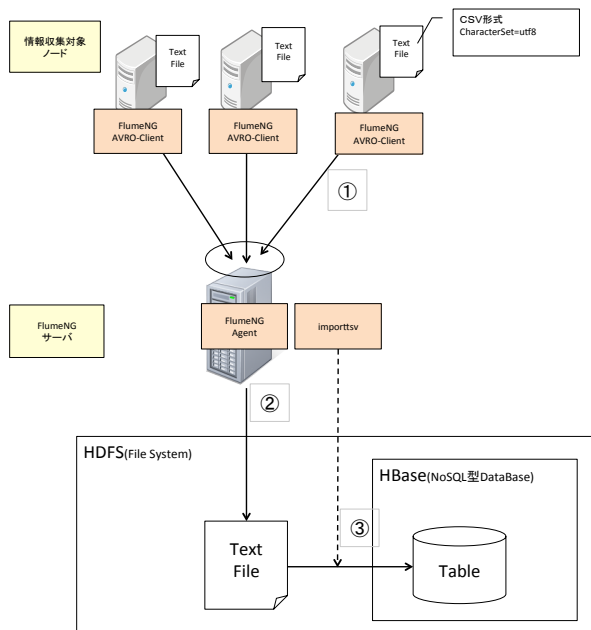
一時的な通信途絶による伝搬遅延が生じた際に、さらに、より長期的な通信途絶が生じた際に、動的な代替通信経路切替え機能を提供するために、ROS の通信基盤に Bluetooth などの短距離無線通信による P2P (Peer-to-Peer) 型伝送機能の追加を行っている。具体的には、ROS と PIAX (P2P Interactive Agent eXtensions - <http://www.piax.org/>) との機能融合を進めている。PIAX が提供する DTN (Delay Tolerant Network) 機能により、P2P フラッディングを行うオーバーレイネットワーク自動構成と、蓄積転送型メッセージ配信が可能となる。これにより、例えば、あるロボットが Wi-Fi 到達性を失った場合にも、その周辺に派遣されている複数のロボットが短距離無線通信範囲を相互に補完しながら P2P オーバーレイネットワークを自動的に構成し、通信到達性を回復することができる。さらに、通信不達の際に転送できなかったデータも、通信到達性が復帰した時点で P2P 層から自動的に送付される。

(2-3) 時系列データ蓄積

ロボットから送付される経時的な環境情報を蓄積し、並行的な解析を可能とするために、大阪大学サイバーメディアセンターで 10 月から稼働を開始したクラスタ型汎用コンピュータシステムに Hadoop (<http://hadoop.apache.org/>) 環境を整備し、外部ソースとのデータ連携方式の研究開発を進めている。

大阪大学クラスタ型汎用コンピュータシステムは、ベース OS として Windows 7、ハイパーバイザ

として Parallels Desktop Workstation Extreme を採用している。Hadoop 環境は、仮想 OS 実行環境の CentOS 上に構築されている。Hadoop 分散ファイルシステムは、各計算ノードが有する SSD 上に構築されている。総容量は 60TB 以上で、仮想 OS 実行環境においてもノードあたり数百 MB/s のデータ入出力が可能である。時系列データを効率的に HDFS に格納するために、HBase (http://hbase.apache.org/) による分散 KVS (Key-Value Store) を構成し、さらに、HDFS へのログ収集機能を提供する Flume (http://flume.apache.org/) を介して、外部ソースから時系列データを投入できる環境を構成している。さらに、ロボットから時系列データを継続的に投入するフロントエンドサービスを開発しており、OpenTSDB (Time-Series Database - http://opentsdb.net/) を組み合わせることで、スケーラブルなデータ蓄積から逐次的なグラフ表示まで可能であることを検証している。



一方、ROS で制御されたロボット側に WebSocket サービスブリッジを組み込むことにより、ロボットが取得した観測データを継続的にウェブサービスに登録するゲートウェイを開発している。Hadoop の認証基盤である Kerberos との認証ブリッジ方式の検討を進めており、認証と認可を伴う経時データ蓄積基盤を構築する予定である。これ

らを連携させることによって得られるプロトタイプシステムが、本グループが目標としているロボット計測システム技術としての適応性を有するか検証を進めている。

(2-4) ROS のクラウド化

これまでの検討を踏まえ、より汎用的な ROS のクラウド対応化手法の検討を行った。通常、ROS で運用されるロボットは、センサーからの情報を定期的に取得し、それをプロセッサにて解析し、アクチュエータを制御している。ROS は、効率的なロボット開発を支援するために、これらのモジュール間の通信をストレージに随時記録し、それを再生して事象を再現する機能 (“ROS Bag” と呼ばれる) を有する。この枠組みをクラウド・ストレージに拡張し、センサーからの情報を全てクラウド・ストレージに格納し、プロセッサはその記録を後追い再生して解析する。解析した制御情報は、同様にクラウド・ストレージを介してアクチュエータに受け渡す (図 6)。この概念設計を元に、要素技術の適否検証を行ない、肯定的な評価を得ている。

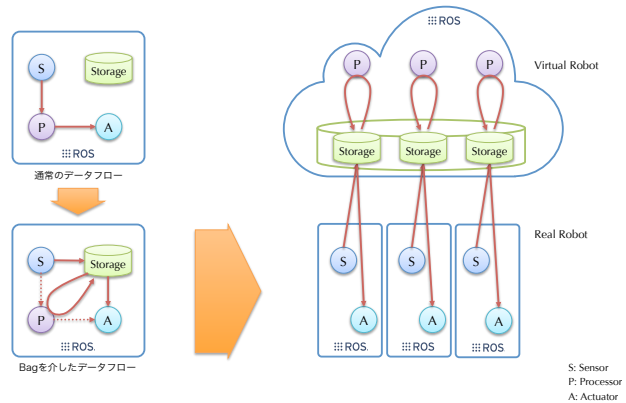


図 6: ROS のクラウドロボット適用

(2) 当初計画の達成状況について

(2-1) 空気冷却風速場のシミュレーション解析と評価

IT コア棟という、設計段階の大規模計算機室を解析対象にして、空調機の位置や台数などの運転条件を現実の機器の動作を想定し変化させながら、空気冷却風速場がある定常状態から別の定常状態

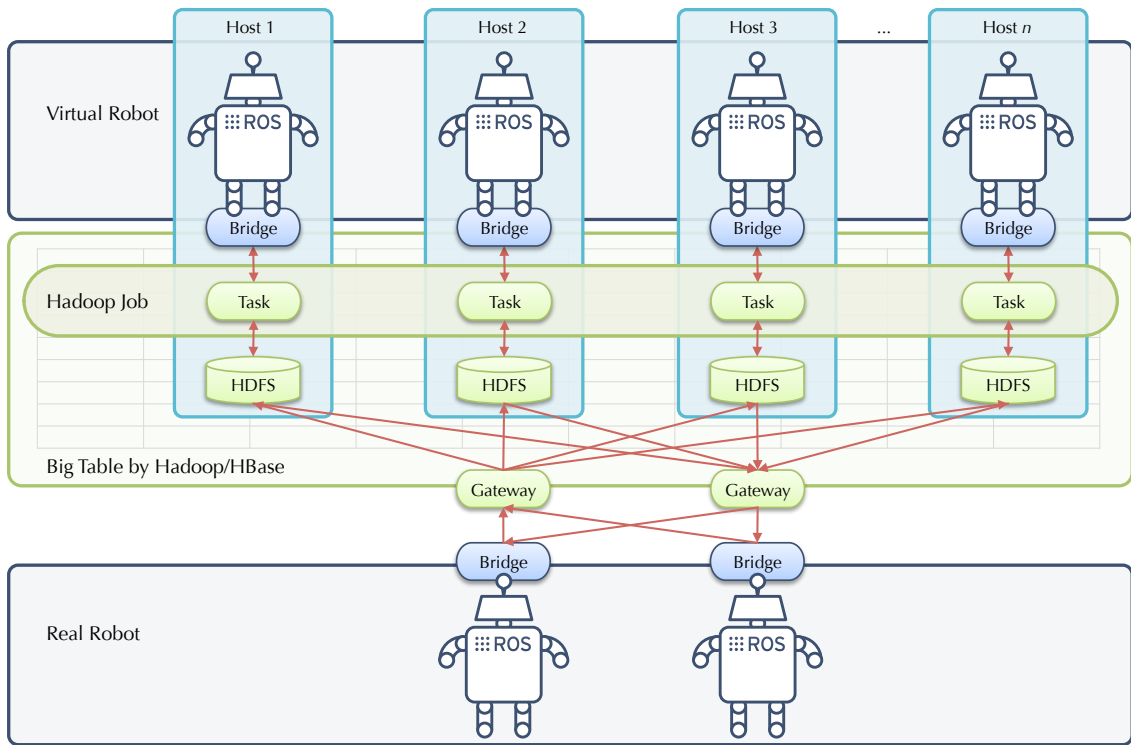


図 7: Hadoop/HBase を介したネットワーク・ロボット連携

に移行する場合のシミュレーション計算を多数実行した。

(2-2) ネットワーク制御ロボットによる巡回監視と連動する計測融合オペレーションシステムの構築

ネットワーク制御ロボットによる巡回環境計測と時系列データの蓄積のための要素技術を網羅的に検証し、個々の要素技術を組み合わせるための連携技術の研究開発を行った。特に、複数のネットワーク制御ロボットを計算機室に巡回させ、取得したセンサー情報を Hadoop 上で解析するための要素技術の開発を行った。

4. 今後の展望

これまでの取組を発展させ、クラウドとロボットの包括的な連携技術として、図 7 に示す iRooHa (inter-ROS Organization on Hadoop/HBase) と呼ぶフレームワーク技術を開発している。これにより、計算機室内を巡回する実ロボットからの環境観測情報を随時、HDFS (Hadoop Distributed Filesystem) にビッグテーブルとして格納し、それを ROS のモジュールとして実装した推論エンジ

ンが解析し、別途格納されている事前シミュレーション結果と比較照合することが可能になると期待している。

さらに、IT コア棟には、フェイルセーフを維持しつつ、これらの推論情報を元に、冷水ポンプや空気循環ファンを制御する機能を導入することを検討している。これに際しては、スーパーコンピュータのジョブスケジューラとの連動も検討しており、図 8 に示すように、IT コア棟自体をクラウドロボットに見立てることも可能であると考えている。

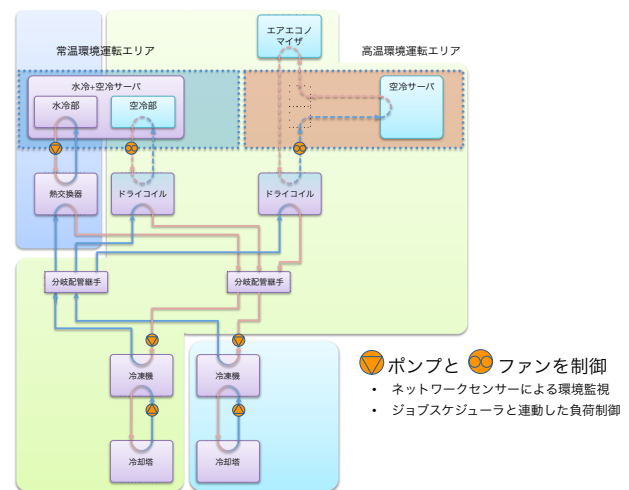


図 8: データセンターのクラウドロボット化

今後、これまで本プロジェクトで利用していた格子ガス法によるシミュレーションや OpenFOAM ツールボックス用に本年度作成したリファレンス・モデルを変換し、結果の照合と精度向上のための施策の検討を行う。また、最終的には、スーパーコンピュータの導入後に環境観測を行い、その結果と照合することによって各手法の優位性の検討を行う。それに際しては、シミュレーション結果をクラウド・ストレージに格納し、機械学習・推論を行って環境観測結果との整合性を採点する手法を確立し、最終的には、実運用における効率向上に結びつけたい。

5. 研究成果リスト

(1) 学術論文

- 合田 憲人, 東田 学, 坂根 栄作, 天野 浩文, 小林 克志, 棟朝 雅晴, 江川 隆輔, 建部 修見, 鴨志田 良和, 滝澤 真一郎, 永井 亨, 岩下 武史, 石川 裕, “高性能分散計算環境のための認証基盤の設計,” 情報処理学会論文誌コンピューティングシステム (ACS) 5(5) pp. 90-102, 2012 年 10 月.

(2) 国際会議プロシーディングス

(3) 国際会議発表

(4) 国内会議発表

- 東田 学, “大規模計算機空気冷却風速場の高解像度解析と適応的クラウドロボット技術による実効的な計測融合オペレーション”, 平成 24 年度「京」を中核とする HPCI システム利用研究課題 中間報告会 ポスター発表, 2013 年 3 月.
- 東田 学, “第二世代クラスタ型汎用コンピュータシステムの紹介”, 第 34 回全国共同利用情報基盤センター研究開発連合発表講演会, 2012 年 11 月.
- 東田 学, “大規模計算機空気冷却風速場の実時間解析と移動型ネットワークセンサー連携による計測融合オペレーション”, 学際大規模情報基盤共同利用・共同研究拠点第 4 回シンポジウム ポスター発表, 2012 年 7 月.

(5) その他 (特許, プレス発表, 著書等)

- 下條 真司, 東田 学, 寺西 裕一, “クラウドロボットから見たネットワークの未来,” 電子情報通信学会誌, 小特集 クラウドネットワークロボット-クラウド化で広がる新たなネットワークロボットの可能性-, vol. 95, no. 12, pp. 1052-1056, 2012 年 12 月.
- 豊川 直子, “CNR (Cloud Network Robot)における通信途絶耐性に関する研究”, 大阪大学工学部, 特別研究報告, February 2013.