

サーバの温度情報に基づくデータセンタ内 リクエスト分散手法の提案

中條 悠介[†] 西 宏章[‡]

[†] 慶應義塾大学理工学部システムデザイン工学科 〒223-8522 神奈川県横浜市港区日吉 3-14-1

E-mail: [†] nakajo@west.sd.keio.ac.jp, [‡] west@sd.keio.ac.jp

あらまし インターネットの普及に伴い、データセンタにおいて処理されるリクエスト量が年々増加している。ロードバランサに代表される負荷分散手法を用いることで、効率的にリクエストを処理できるようになるが、データセンタにおけるサーバから発生する熱を考慮できていない。リクエスト処理に伴うサーバなどの IT 機器の異常な発熱は、空調機器の消費電力や IT 機器の故障率を増大させ、運用コストの増加へとつながる。そこで本報告では、到着したリクエストの処理負荷分散を行いながら運用コストを削減するべく、サーバの温度情報を考慮したリクエスト分散手法を提案する。データセンタ熱モデルを設計し、実際のリクエストデータを用いてシミュレーションにより提案手法の評価を行う。

キーワード データセンタ, ロードバランシング, 熱モデルシミュレーション

Request Distribution Based on Server-Temperature in Data Center

Yusuke NAKAJO[†] and Hiroaki NISHI[‡]

[†] Department of System Design Engineering 3-14-1 Hiyoshi, Yokohama-shi Kanagawa, 223-8522 Japan

E-mail: [†] nakajo@west.sd.keio.ac.jp, [‡] west@sd.keio.ac.jp

Abstract With the growth of the Internet, data centers have to deal with the increase in the amount of data. Load Balancing technologies can help data centers process data effectively; however it does not take into account the heat generated by servers. Excessive heat could increase the failure rate of IT devices, which will lead to high maintenance costs in data centers. In this report, we propose a novel request distribution system based on server-temperature to reduce the maintenance cost in data centers and evaluate the proposed system by modeling a thermal model of a data center.

Keywords Data Center, Load Balancing, Thermal Model Simulation

1. はじめに

近年、スマートフォンの台頭や IoT(Internet of Things)によるサービス多様化,そしてクラウドコンピューティングサービスの出現により、インターネット社会は急速に発展している。それに伴い、ネットワークトラフィック量は爆発的に増加し、そのデータを処理するデータセンタの需要が高まっている。また、このようなデータセンタの需要の増加に伴い、国内では大規模なデータセンタの新設、増設が相次いでおり、データセンタサービス市場における競争が激化している。IDC Japan の調査によると国内のデータセンタサービス市場は年々増加しており、2015 年から 2020 年までの年間平均成長率(Compound Annual Growth rate, CAGR)は 7.0%と順調に成長すると予測している[1]。このような状況下においてデータセンタ事業者には効率的なデータセンタの運用が求められる。データセン

タの運用コストには主に消費電力と、IT 機器の故障や寿命による取り換えといったライフサイクルコストが挙げられる。一般的なデータセンタにおいて、IT 機器の消費電力はデータセンタの総消費電力の 36%を占めており、残りの内訳は、空調機器が 50%, UPS 発生損失, バッテリの充電が 18%, 照明およびその他が 3%である[2]。IT 機器と空調機器の消費電力が総消費電力の 86%であることから、データセンタの運用コストを削減するうえで、IT 機器および空調機器の消費電力を削減することは重要な課題であるといえる。

本報告では、空調機器の消費電力削減を行うために、サーバの排気温度の最大値を低減させることに注目した。サーバの排気温度の最大値を低減させることで空調機器の設定温度を高く設定することができ、冷却効率を高く維持することができる。これにより冷却における余分な消費電力を削減することが可能である[3]。

データセンタにおけるライフサイクルコストは、サーバが発生する熱と関係している。リクエスト処理によるサーバの発熱は、サーバやその他 IT 機器の動作環境の温度増加を招き、故障率の増加や寿命の低下へとつながる。例えば、半導体および電子部品は熱によって故障率が大きくなり、寿命が縮まることが確認されている。

本報告では、サーバにおいて多く用いられている半導体の故障率に着目する。故障原因の多くは化学反応に依る物質の変質にあり、高温環境下において促進される。半導体の温度ストレスに依る故障の依存性はアレニウスモデルで表現できることが知られており[4]、この反応速度論モデルを用いて評価を行う。

本報告では、データセンタにおける空調の消費電力削減および IT 機器の故障率を低減するために、データセンタ内のサーバの温度情報を基に、データセンタへのリクエストを各サーバに分配する機構を提案する。

2. 関連研究

データセンタの空調システムにおけるエネルギー消費量および運用コストの削減に関する研究が盛んに行われている。Sullivan らは、冷気を供給するサーバラックの吸気口側をコールドアイル、排気口側をホットアイルとして、その二つの気層を分離する配置方法を提案した[5]。この手法では、二列のサーバラックの吸気口が向かいあうように配置されている。現在多くのデータセンタではホットアイルおよびコールドアイルを設けている。

沼田らは、この考え方を発展させ、4 点について対策を行うことで、サーバラックの吸気温度を低下させ、空調の冷却効率を改善できることを示している[3]。1 つめは、サーバラックの向きと床からの冷気の取り入れ口の配置を換えて、通路に冷気通路と暖気通路を設置すること。2 つめは、隣り合うサーバラック同士の間隔をなくし風の逆流を防止すること。3 つめは、サーバラック内の空きスペースにブランクパネルを取り付け、風の逆流を防止すること。4 つめは、サーバラックの端に仕切り板を取り付け、冷気が空調機器に直接戻るのを防止すること。結果的に、サーバラックの吸気温度の最高値は改善前の 33.9℃から改善後の 27.9℃に下げることが可能であると示した。

以上のように、熱設計モデルからデータセンタ内のラックの構成の仕方を工夫することで、消費電力の削減を図る研究がある一方で、IT 機器の消費電力について着目し、リクエストの処理に必要なサーバのみを稼働することでアイドル電力を削減する方法に関する研究も行われている。Ahmad らはアイドル電力と冷却電力はトレードオフの関係にあることに注目し、このトレードオフの関係を最適化させることで、総消費電力

の削減を図った[6]。少数のサーバに負荷を集中させることによって、アイドル電力を削減する手法は Spatial Subsetting と呼ばれ、反対に多数のサーバに負荷を分散させることでサーバの温度上昇を防ぎ、冷却電力を削減する手法は Inverse-Temperature と呼ばれる。

Spatial Subsetting 法ではサーバのアイドル電力は削減できるが、少数のサーバに熱が集中し、そのサーバを冷却するために空調機器の電力が大きくなる。一方 Inverse-Temperature 法では、サーバの負荷を分散させることでサーバの発生熱量を抑え、空調機器の消費電力を削減することができるが、多くのサーバを稼働させる必要があり、IT 機器の消費電力が増加する。

また、Ahmad[7]らは、サーバの温度によってサーバを Hot Zone, Warm Zone, Cool Zone の三種類に分類し、場所ごとに Spatial Subsetting 法と Inverse-Temperature 法を使い分ける手法を提案した。結果として、総消費電力を最大で 23%削減できることを示した。

本報告では、これら既存研究とは異なり、データセンタにおけるロードバランシングに着目した。データセンタにおいて負荷分散方式は安定したリクエストの処理を可能にする。ロードバランシングの方式には様々な種類が存在するが、データセンタの効率的運用を目的としたロードバランシングは存在しない。例えば、代表的な負荷分散技術であるラウンドロビンではリクエストを負荷分散対象のサーバに順番に振り分けるが、大きなサイズのリクエストが連続的に同じサーバに割当てられる可能性があり、サーバが高温状態になると考えられる。サーバの高温状態は故障率の増加および冷却コストの増加へとつながるため、サーバの温度情報を考慮した負荷分散手法の提案を行うことで、データセンタの運用コスト削減を達成する。

3. モデル設計

提案手法の評価に用いたシミュレーションモデルの構築方法について、サーバ、ラック、データセンタのモデル構成を順に説明する。まず、ラック内のサーバモデルの作成方法について説明する。図 1 はサーバモデルの模式図を表す。

コールドアイル側から空調機器による冷却風がサーバへと届き、サーバ内を通過し、ホットアイル側へと排気される。サーバ間のエネルギー保存の関係から以下の式(1)が成立する。



図 1 サーバモデルの模式図

$$q_{IT} = M_s \cdot c_{p,s} \cdot \frac{dT_s}{dt} + \dot{m} \cdot c_{p,air}(T_{ex} - T_{in}) \quad (1)$$

つぎに、サーバ間を通る流体とサーバ固体間でやりとりされる熱量についてもエネルギー保存の法則より、以下の式 (2) が成立する。

$$M_s \cdot c_{p,s} \cdot \frac{dT_s}{dt} = hA(T_{in} + \Delta T_{IT} - T_s) \quad (2)$$

q_{IT} はサーバの発熱量、 $M_s \cdot c_{p,s}$ はサーバの熱容量、 T_s はサーバの温度、 $\dot{m}$ はサーバ間を通る空気の流れ、 $c_{p,air}$ は空気の比熱、 T_{ex} は排気温度、 T_{in} は吸気温度、 h は熱伝達率、 A はサーバの表面積、 ΔT_{IT} はサーバ内発熱による温度変化を表す。コールドアイルからサーバへ流入する単位時間当たりの熱量は空調機器からの冷却風量によって決定され、風量は常に一定であると仮定してシミュレーションを行う。

式(2)における熱伝達率 h 及びサーバ内部の流体と熱交換を行うサーバの表面積 A を推定することは困難である。そこで本報告では、VanGilderらのモデルの作成方法に従い、サーバ固体とサーバ固体間を流れる流体との間でやり取りされる熱量を表す無次元量 ε を導入した。この結果、計測や導出が困難な式(2)の h, A を直接的に求めることを回避できる。VanGilderらによれば ε は次の式(3)のように定義されている[8]。

$$\varepsilon = \frac{T_{in} - T_{ex}}{T_{in} - T_s} = \frac{hA}{\dot{m}c_{p,air}} \quad (3)$$

この式(3)を用いることで、式(2)は式(4)のように書き直すことができる。 ε は、VanGilderらが実験的に測定した値を用いた。 ε はサーバの密度と密接な関係があり、Pardeyらは実験的にサーバの密度とサーバの熱効率 ε の関係を求めている[9]。

$$M_s \cdot c_{p,s} \cdot \frac{dT_s}{dt} = \varepsilon \dot{m} c_{p,air}(T_{in} + \Delta T_{IT} - T_s) \quad (4)$$

本報告におけるシミュレーションでは、2U-Dot Hill, DH0149-01 をサーバとして用いた。Pardeyらの研究では、2U-Dot Hill, DH0149-01 の ε の値に関して、風量に依存せず、0.97 であると結論づけている。そこで、本報告では ε の値を 0.97 であるとしてサーバモデルの作成を行うことにした。なお、サーバにあたる冷却風量は 300(cfm)で一定とした。

続いて、ラック内におけるラックマウント型サーバを単一ラックモデルへと拡張する。単一ラック内において n 個のサーバが収納されることを想定し、各サーバは前述したようにモデル化を行なう。単一ラック内におけるサーバの熱容量を式(5)に示す。 $M_{RCp,R}$ はラックの熱容量を表す。単一ラック内における発熱量 Q_{rack} を式(6)に示す。

$$M_{total}c_{p,total} = M_{RCp,R} + \sum_{i=1}^n M_s c_{p,s} \quad (5)$$

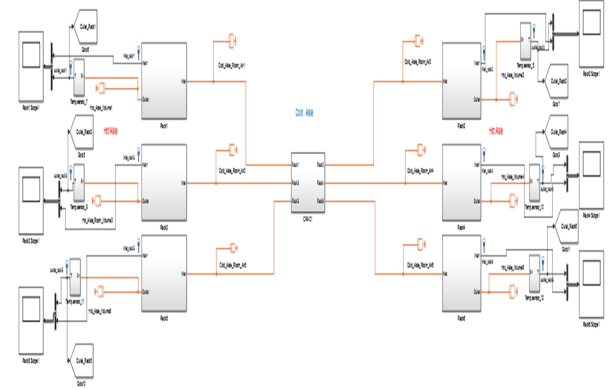


図 2 データセンタモデル

$$Q_{rack} = P_{base} + \sum_{i=1}^n q_{IT} \\ = P_{base} + \sum_{i=1}^n (P_{idle} + k u) \quad (6)$$

P_{base} (ベース消費電力量) は、ラックが消費する電力量を表す。ラックに直接搭載される冷却モジュールや電源モジュールなどの機器による、アイドル状態においても消費される電力を示す。サーバの消費電力はアイドル状態における消費電力と CPU 使用率によって変動する消費電力の和として表される。

最後にモデル化を行ったラックを、コールドアイル、ホットアイルを考慮して配置し、簡易的なデータセンタモデルを作成した。コールドアイル側の冷却風とホットアイル側の暖気の混合を防ぐために、アイルキャッピング型のラック構成を想定した。図 2 は実際に作成したデータセンタモデルであり、このモデルを使用して提案手法の評価を行った。

以下に本シミュレーションのモデル設計における前提条件を記述する。

- Samiら[10]の先行研究に基づき、ラックの熱容量に関しては無視する。
- ラック内において温度分布は生じない。
- Pardeyら[9]の先行研究に基づき、サーバ内における発熱は中心部分から一様に消散する。
- 排気口からの暖気は逆流してコールドアイル側へと循環しない。
- コールドアイル側は常に設定温度を保ち、ラックの吸気口側には常に設定温度の冷風が当たる。
- データセンタ内部のみを考慮しており、外部からの熱の影響はない。

4. 提案手法

ロードバランサなどの負荷分散手法を用いることで、データセンタにおいて効率的にリクエストを処理することが可能である。現在、様々なロードバランシング手法が提案されており、目的によって使い分けが

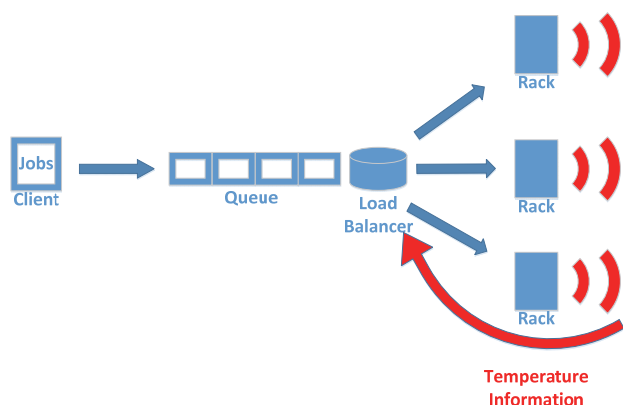


図 3 提案機構の概要

なされている。しかし、これらの負荷分散手法は、データセンタの運用コストを考慮しておらず、ただ安定的に処理を行うことを目的としている。

そこで本報告では、データセンタ内のラックの排気温度情報を考慮し、クライアントからのリクエストを各ラックに分配するロードバランサの提案を行う。本提案によれば、負荷分散による安定的な処理を実現しつつ、各ラックの排気温度を低減することができる。従って、空調の消費電力削減、および IT 機器の故障率の低減を実現できる。図 3 に提案機構の概要を示す。データセンタ内のロードバランサがラックの温度情報を考慮しながら、リクエストが到着する毎に温度の低いラックに対して優先的にリクエストの分配を行う。

5. 評価

本報告では、データセンタに送られるリクエスト量として、2015 年 6 月 11 日に幕張メッセにて開催された Interop Tokyo 内でやりとりされた HTTP パケットの送受信データをシミュレーションに用いた。図 4 に使用したリクエストデータを示す。データセンタのモデル作成には、数値解析ソフトウェア MATLAB とモデ

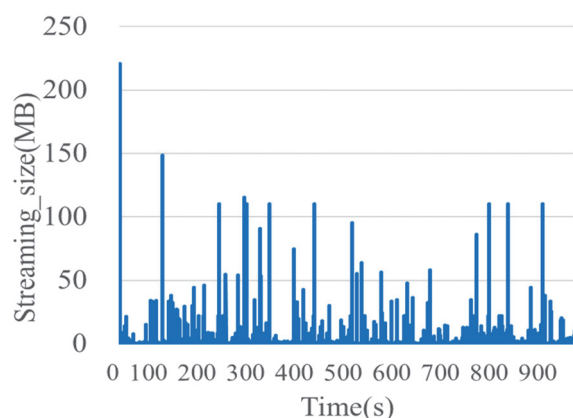


図 4 Interop2015 における HTTP リクエストデータ

表 1 シミュレーションに用いたパラメータ

パラメータ	値
シミュレーション時間	900(s)
使用したサーバ	2U-Dot Hill, DH0149-01
サーバの比熱	460(J/kg · K)
ラック内のサーバの数	22
空気の比熱	1,005(J/kg · K)
空気の密度	1.247(kg/m ³)
コールドアイルの設定温度	10(°C)
コールドアイル側からの風量	300(cfm)
ラックの数	6

リングおよびシミュレーションのためのダイナミックシステム Simulink を用いた。

シミュレーションを行う上で、データセンタの熱モデルを基にロードバランサによってリクエストの分散を行うことから、電気、機械、情報、熱の分野を横断してシミュレーションを行う必要がある。こういった背景から分野横断的に評価をとることのできる MATLAB および Simulink が最も適していると判断した。

提案手法と 2 種類のロードバランシング手法との比較評価を行った。1 つめは、到着したリクエストに対して、処理を行うサーバを順番に決定してリクエスト分配する方法であるラウンドロビン、2 つめは、到着したリクエストに対して、最も CPU 使用率が低いサーバに対して、リクエストを分配する方法である CPU 負荷監視による動的負荷分散手法である。シミュレーシ

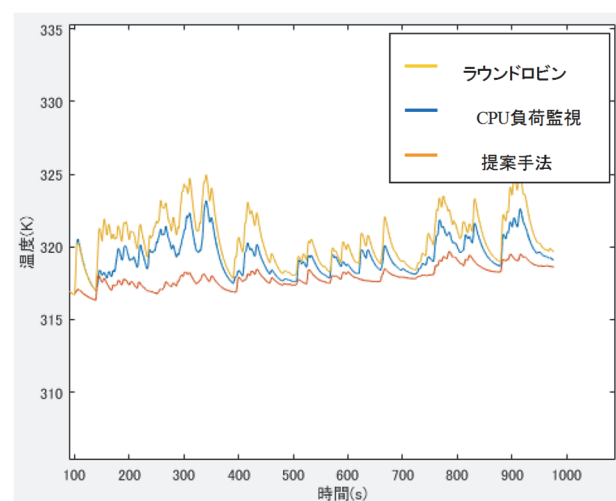


図 5 6つのラックの排気温度の平均値

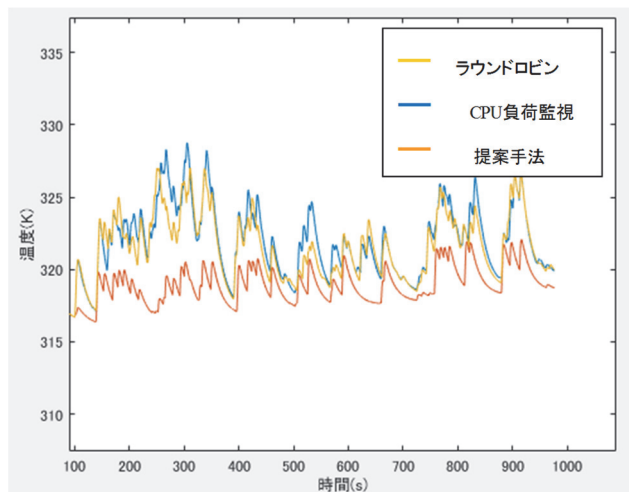


図 6 6つのラックの排気温度の最大値

オン時間は 977s とし、開始から経過時間 100s のところで、データセンタモデルに対してリクエストの送信を行った。シミュレーションにおいて用いたパラメータとそれぞれの値を表 1 に示す。

全ラックの排気温度の平均値及び最大値はそれぞれ図 5 と図 6 に示す。いずれの図も縦軸は温度を表し、横軸はシミュレーション時間を表す。また、黄色の線はラウンドロビン、青色は CPU 負荷監視による動的負荷分散、橙色は提案手法の温度変化を表している。図 5 と図 6 より、提案手法の温度変化は他の 2 手法と比べると最も安定かつ低い値を示すことがわかる。これは提案手法が全ラック間において温度差が縮まるようにリクエスト分散していることを示している。また既存手法において、ラウンドロビンよりも CPU 監視による動的負荷分散手法の方が平均値が低い。これは、CPU 使用率を利用する場合、比較的溫度上昇の小さい

表 2 排気温度の平均値および最大値の比較

	ラウンドロビン	CPU負荷監視	提案手法
平均温度(K)	320.72	319.45	317.92
最高温度(K)	327.04	328.79	322.09

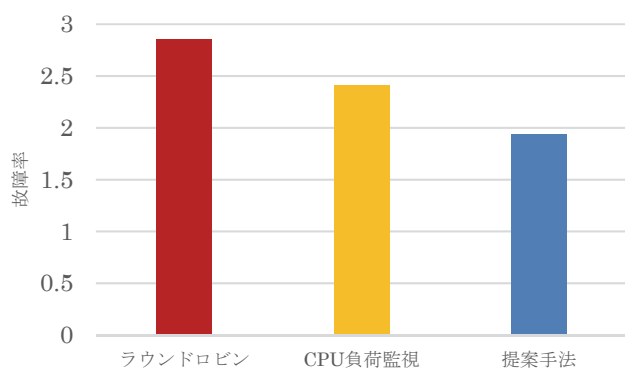


図 7 各手法における故障率の比較

ラックへリクエストが送られているためである。しかし、提案手法と比べると、CPU 負荷監視による動的負荷分散は排気温度の最大値および平均値が高い。これは、CPU 使用率の変動時間と温度変化の時定数の違いが原因であると考えられる。CPU 使用率の変動時間は温度変化の時定数と比べると短い。そのため、CPU 使用率が低下しても、温度が低下するまでは時間がかかってしまう。つまり、CPU 使用率が低いサーバに対してリクエストを送ったとしても、温度が高いサーバに送っている可能性がある。提案手法は、温度が低いサーバに対してリクエストの分配を行うので、動的負荷分散手法と比較して、ラックの排気平均温度および排気最大温度に差が生じたと考えられる。提案手法と 2 つの既存手法における 6 つのラックの排気温度の平均値及び最大値を表 2 にまとめた。

平均温度に関して、提案手法ではラウンドロビンと比較して、最大で約 2.80(K)温度が低下した。また、最高温度に関しては CPU 負荷監視手法と比較すると、提案手法では最大で約 6.70(K)温度が低下した。また、故障率に関して、一般的な IT 機器の動作適正温度を 40(°C)としたときの半導体の故障率をアレニウスの式より求めた。各手法における故障率の比較を図 7 に示した。図 7 より、提案手法においてはラウンドロビンと比較した場合、故障率が約 32%低減し、一方 CPU 負荷監視と比べた場合、故障率が約 19%低減したことが確認できる。

6. 結論と今後の課題

インターネット社会の急速な発展に伴い、データセンタにおいて処理されるデータの数が年々増加し、データセンタの需要が高まっている。このような状況に伴い、データセンタ事業者は効率的なデータセンタ運用が求められている。本報告では、データセンタにおける空調機器の消費電力の削減および IT 機器のライフサイクルコストを低下させるために、温度情報を考慮したデータセンタ内のリクエスト分散手法を提案した。また簡易的なデータセンタモデルを作成しシミュレーションを行うことで、提案手法と既存 2 手法との比較を行った。結果として、平均温度に関してはラウンドロビンと比較すると最大で約 2.80(K)温度が低下し、最高温度に関しては CPU 負荷監視と比較すると提案手法では、最大で約 6.70(K)温度が低下した。また、故障率に関してはラウンドロビンと比較すると約 32%、CPU 負荷監視と比較すると約 19%低減できると確認できた。

今後の課題として、シミュレーションモデル構築法の妥当性の検証、提案アルゴリズムによる遅延の評価、ラック内の厳密化の 3 つが挙げられる。まずシミュレーションモデルの妥当性について、本来ならばラック

内における発熱の影響がどのように排気口温度に影響するかを実機によって検証し、作成したモデルとの比較が必要であるため、実機環境でのシミュレーションモデルの検証が課題として挙げられる。次に、提案手法の遅延時間を評価する必要がある。データセンタ運用効率を考えるあまり処理時間が遅くなるとは本末転倒である。他の既存手法と比べてどの程度の処理遅延が提案アルゴリズムによって生じるかといった検証が課題として挙げられる。ラック内の厳密化であるが、本報告の提案手法により行ったラック単位のタスク分散ではなく、ラック内のサーバ温度に基づいてサーバ単位でタスク分散を行うことが課題として挙げられる。実際にはラック内において、各サーバにあたる冷却風量には違いが生じることを考慮して、ラック内のサーバ単位での負荷分散を行うとより効果的である可能性がある。

謝 辞

本報告は、公益財団法人セコム科学技術振興財団助成、文部科学省科学技術研究費基盤研究B (JP16H04455)、平成28年度総務省委託研究開発「スマートコミュニティサービス向け情報通信プラットフォームの研究開発」、国土交通省住宅・建築物技術高度化事業ならびに慶應義塾大学グローバルスマート社会創造プロジェクト研究の一環としてなされた。

文 献

- [1] “国内データセンターサービス市場予測を発表.”
[Online]. Available:
<http://www.idc-japan.co.jp/Press/Current/20161207Apr.html>. [Accessed: 04-Jan-2017].
- [2] リチャード L, “データセンタの総所要 電力の算出,” *APC Legend. Reliab.*, vol. White Paper, no. #3.
- [3] K. Numata, H. Okamura, H. Sasaki, T. Nomura, and K. Words, “Energy Efficiency Simulation Model of Data Centers,” no. 56, pp. 65–70, 2008.
- [4] “温度による加速 | 東芝 ストレージ&デバイスソリューション社.” [Online]. Available:
<http://toshiba.semicon-storage.com/jp/design-support/reliability/device/testing/testing2/1271185.html>. [Accessed: 21-Jan-2017].
- [5] B. F. Robert, “Alternating Cold and Hot Aisles Provides More Reliable Cooling for Server Farms.”
- [6] F. Ahmad and T. N. Vijaykumar, “Joint Optimization of Idle and Cooling Power in Data Centers While Maintaining Response Time.”
- [7] F. Ahmad and T. N. Vijaykumar, “Joint optimization of idle and cooling power in data centers while maintaining response time,” *ACM SIGPLAN Not.*, vol. 45, no. 3, p. 243, 2010.
- [8] J. W. VanGilder, C. M. Healey, Z. M. Pardey, and X. Zhang, “A compact server model for transient data center simulations,” vol. 119, pp. 358–370, 2013.
- [9] Z. M. Pardey and J. W. VanGilder, “Further exploration of a compact transient server model,” in *Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2014, pp. 1322–1329.
- [10] S. Alkharabsheh, B. Sammakia, S. Shrivastava, and R. Schmidt, “Dynamic models for server rack and CRAH in a room level CFD model of a data center,” in *Fourteenth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2014, pp. 1338–1345.